

PCT

09/530-233
WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



#3
CO: 1

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶: C07K 14/47, 16/44, C12N 15/12, 15/63, 1/21, 5/10	A1	(11) International Publication Number: WO 97/01577 (43) International Publication Date: 16 January 1997 (16.01.97)
(21) International Application Number: PCT/GB96/01523 (22) International Filing Date: 25 June 1996 (25.06.96) (30) Priority Data: 9513180.1 28 June 1995 (28.06.95) GB (71) Applicant: UNIVERSITY COLLEGE LONDON [GB/GB]; Gower Street, London WC1E 6BT (GB). (72) Inventors: WOOD, John, Nicholas; Gower Street, London WC1E 6BT (GB). AKOPIAN, Armen, Norakovitch; Gower Street, London WC1E 6BT (GB). (74) Agent: TINSLEY, Rachel, Maria; Zeneca Pharmaceuticals, In- tellectual Property Dept., Mereside, Alderley Park, Maccles- field, Cheshire SK10 4TG (GB).		(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the</i> <i>claims and to be republished in the event of the receipt of</i> <i>amendments.</i>
(54) Title: ION CHANNEL (57) Abstract The present invention relates to a novel 1,957 amino acid tetrodotoxin-insensitive voltage-gated sodium channel specifically located in mammalian sensory neurons. Nucleic acid sequences coding for the novel sodium channel, vectors, host cells and methods of identifying modulators of the novel sodium channel for use in treatment of pain are also provided.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM Armenia
AT Austria
AU Australia
BB Barbados
BE Belgium
BF Burkina Faso
BG Bulgaria
BJ Benin
BR Brazil
BY Belarus
CA Canada
CF Central African Republic
CG Congo
CH Switzerland
CI Côte d'Ivoire
CM Cameroon
CN China
CS Czechoslovakia
CZ Czech Republic
DE Germany
DK Denmark
EE Estonia
ES Spain
FI Finland
FR France
GA Gabon

GB United Kingdom
GE Georgia
GN Guinea
GR Greece
HU Hungary
IE Ireland
IT Italy
JP Japan
KE Kenya
KG Kyrgyzstan
KP Democratic People's Republic
of Korea
KR Republic of Korea
KZ Kazakhstan
LI Liechtenstein
LK Sri Lanka
LR Liberia
LT Lithuania
LU Luxembourg
LV Latvia
MC Monaco
MD Republic of Moldova
MG Madagascar
ML Mali
MN Mongolia
MR Mauritania

MW Malawi
MX Mexico
NE Niger
NL Netherlands
NO Norway
NZ New Zealand
PL Poland
PT Portugal
RO Romania
RU Russian Federation
SD Sudan
SE Sweden
SG Singapore
SI Slovenia
SK Slovakia
SN Senegal
SZ Swaziland
TD Chad
TG Togo
TJ Tajikistan
TT Trinidad and Tobago
UA Ukraine
UG Uganda
US United States of America
UZ Uzbekistan
VN Viet Nam

ION CHANNEL

Voltage-gated sodium channels are transmembrane proteins which cause sodium permeability to increase. Depolarization of the plasma membrane causes sodium channels to open allowing sodium ions to enter along the electrochemical gradient creating an action potential.

Voltage-gated sodium channels are expressed by all electrically excitable cells, where they play an essential role in action potential propagation. They comprise a major subunit of about 2000 amino acids which is divided into four domains (D1-D4), each of which contains 6 membrane-spanning regions (S1-S6). The alpha-subunit is usually associated with 2 smaller subunits (beta-1 and beta-2) that influence the gating kinetics of the channel. These channels show remarkable ion selectivity, with little permeability to other monovalent or divalent cations. Patch-clamp studies have shown that depolarisation leads to activation with a typical conductance of about 20pS, reflecting ion movement at the rate of 10^7 ions/second/channel. The channel inactivates within milliseconds (Caterall, W.A., *Physiol. Rev.* 72, S4-S47 (1992); Omri et al, *J. Membrane Biol* 115, 13-29; Hille, B, *Ionic Channels in Excitable Membranes*, Sinauer, Sunderland, MA (1991)).

Sodium channels have been pharmacologically characterised using toxins which bind to distinct sites on sodium channels. The heterocyclic guanidine-based channel blockers tetrodotoxin (TTX) and saxitoxin (STX) bind to a site in the S5-S6 loop, whilst μ -conotoxin binds to an adjacent overlapping region. A number of toxins from sea anemones or scorpions binding at other sites alter the voltage-dependence of activation or inactivation.

Voltage-gated sodium channels that are blocked by nanomolar concentrations of tetrodotoxin are known as tetrodotoxin sensitive sodium channels (Hille (1991) "Ionic Channels in Excitable Membranes", Sinauer Sunderland, MA (1991)) whilst sodium channels that are blocked by concentrations greater than 1 micromolar are known as tetrodotoxin-insensitive (TTXi) sodium channels (Pearce and Duchen *Neuroscience* 63, 1041-1056 (1994)).

Dorsal root ganglion (DRG) neurons express at least three types of sodium channels which differ in kinetics and sensitivity to TTX. Neurons with small-diameter cell bodies and unmyelinated axons (C-fibers) include most of the nociceptor (damage-sensing)

population and express a fast TTX-sensitive current and a slower TTX-insensitive current. Of the five cloned sodium channel α -subunit transcripts known to be present in dorsal root ganglia, none exhibits the properties of the TTX-insensitive channel.

Sodium channel blockers are used clinically to provide pain relief. Three
5 classes of sodium channel blockers in common clinical use are: local anesthetics such as lidocaine, some anticonvulsants such as phenytoin and carbamazepine, and some antiarrhythmics such as mexiletine. Each of these is known to suppress ectopic peripheral nervous system discharge in experimental preparations and to provide relief in a broad range of clinical neuropathic conditions.

10 Applicants have now found a novel voltage-gated sodium channel (hereinafter referred to as a sodium channel specifically located in sensory neurons or also referred to as SNS sodium channel) that is present in sensory neurons (or neurones) but not present in glia, muscle, or the neurons of the sympathetic, parasympathetic, enteric or central nervous systems. Preferably the sodium channel of the invention is found in the
15 neurons of the dorsal root ganglia (DRG) or cranial ganglia. More preferably the sodium channel of the invention is found in the neurons of the dorsal root ganglia. Preferably the sodium channel is specifically located in rat sensory neurons or human sensory neurons.

The sodium channel of the present invention is believed to play a role in nociceptive transmission because some noxious input to the central nervous system is
20 known to be insensitive to TTX. Persistent activation of peripheral nociceptors has been found to result in changes in excitability in the dorsal horn associated with the establishment of chronic pain. Increased sodium channel activity has also been shown to underlie neuroma-induced spontaneous action potential generation. Conversely, chronic pain may be successfully treated by surgical or pharmacological procedures which block
25 peripheral nerve activation. Blockage of nociceptor input may therefore produce useful therapeutic effects, even though central nervous system plasticity plays a pivotal role in the establishment of chronic pain. Sensory neuron-specific voltage-gated sodium channels, particularly sub-types associated with a nociceptive modality such as the sodium channel of the invention, thus provide targets for therapeutic intervention in a range of pain states.
30 The electrophysiological and pharmacological properties of the expressed SNS sodium channel are similar to those described for the small diameter sensory neuron tetrodotoxin-resistant sodium channels. As some noxious input into the spinal cord is resistant to

tetrodotoxin, block of expression or function of such a C-fiber-restricted sodium channel may have a selective analgesic effect.

In another aspect the present invention provides an isolated protein comprising a sodium channel specifically located in rat sensory neurons as encoded by the insert deposited in NCIMB deposit number 40744, which was deposited at The National Collections of Industrial and Marine Bacteria, 23 St Machar Drive, Aberdeen AB2 1RY, Scotland, United Kingdom on 27 June 1995 in accordance with the Budapest Treaty.

The invention also provides nucleotide sequences coding for the SNS sodium channel. In a preferred embodiment, the nucleotide sequence encodes a sodium channel specifically located in rat sensory neurons which is as set out in Figure 1a or a complementary strand thereof.

The approximately 6.5 kilobase (kb) transcript expressed selectively in rat dorsal root ganglia that codes for the novel sodium channel of the invention shows sequence similarities with known voltage-gated sodium channels. The cDNA codes for a 1,957 amino acid protein. In particular, the novel sodium channel of the invention shows 65% identity at the amino acid level with the rat cardiac tetrodotoxin-insensitive (TTXi) sodium channel. The aromatic residue that is involved in high-affinity binding of TTX to the channel atrium of TTX-sensitive sodium channels is altered to a hydrophilic serine in the predicted protein of the SNS sodium channel, whereas the residues implicated in sodium-selective permeability are conserved. The novel sodium channel specifically located in sensory neurons shows relative insensitivity to TTX ($IC_{50} > 1$ micromolar) and thus exhibits properties different from other cloned sodium channel transcripts known to be present in dorsal root ganglia.

The invention also provides expression and cloning vectors comprising a nucleotide sequence as hereinabove defined. In order to effect transformation, DNA sequences containing the desired coding sequence and control sequences in operable linkage (so that hosts transformed with these sequences are capable of producing the encoded proteins) may be included in a vector, however, the relevant DNA may then also be integrated into the host chromosome.

The invention also provides a screening assay for modulators of the sodium channel which is specifically located in sensory neurons wherein the assay comprises

adding a potential modulator to a cell expressing the SNS sodium channel and detecting any change in activity of the sodium channel.

The present invention also provides a modulator which has activity in the screening assay hereinabove defined. Modulators of the sodium channel as hereinabove
 5 defined are useful in modulating the sensation of pain. Blockers of the sodium channel will block or prevent the transmission of impulses along sensory neurons and thereby be useful in the treatment of acute, chronic or neuropathic pain.

The present invention thus relates to novel voltage-gated sodium channel proteins specific to sensory neurons, to nucleotide sequences capable of encoding these
 10 sodium channel proteins, to vectors comprising a nucleotide sequence coding for a sodium channel of the invention, to host cells containing these vectors, to cells transformed with a nucleic acid sequence coding for the sodium channel, to screening assays using the sodium channel proteins and/or host cells, to complementary stands of the DNA sequence which is capable of encoding the sodium channel proteins and to antibodies specific for the sodium
 15 channel proteins. These and other aspects of the present invention are set forth in the following detailed description.

Brief Description of the Drawings:

Figure 1a shows the nucleic acid and amino acid sequences of the sodium
 20 channel specific to the rat DRG (SNS-B) (SEQ ID NO: 1 and SEQ ID NO: 2).

Figure 1b shows the structure of the SNS-B voltage-gated sodium channel in pGEM-3Z.

Figure 1c shows a schematised drawing of a known voltage-gated sodium channel.

Figure 2 shows sequences of examples of PCR primers for isolation of
 25 human clone probes. RLLRVFKLAKSWPTL - SEQ ID NO: 21; 5' gcttgctgcgggtcttcaagc 3' SEQ ID NO: 22; LRALPLRALS RFEG - SEQ ID NO: 23; 5' atcgagacagagcccgagcgcg 3' SEQ ID NO: 24; 5' acgggtgccgcaaggacggcgtctccgtgtggaacggcgagaag 3' SEQ ID NO: 25; and 5' ggctatccttctcttccagctctcaccaggtatggagccaggt 3' - SEQ ID NO: 26.

Figure 3 shows a film of ³⁵S radio-labelled SNS-B voltage-gated sodium
 30 channel protein in a coupled transcription/translation system.

Figure 4a and Figure 4b show SNS-GST fusion protein constructs for antibody generation. TCCCGTACGCTGCAGCTCTTT - SEQ ID NO: 27; CCCGGGGAAGGCTAC - SEQ ID NO: 28; GTCGACACCAGAAAT - SEQ ID NO: 29; GGATCCTCTAGAGTCGACCTGCAGAAGGAA - SEQ ID NO: 30

5

In accordance with one aspect of the invention there is provided an isolated and/or purified nucleic acid sequence (or polynucleotide or nucleotide sequence) which comprises a nucleic acid sequence which encodes the mammalian sodium channel specifically located in sensory neurons or a complementary strand thereof. Preferably, the nucleic acid sequence encodes the sodium channel specifically located in mammalian dorsal root ganglia. More preferably, the nucleic acid sequence encodes the rat or human sodium channel specifically located in dorsal root ganglia. The rat nucleic acid sequence preferably comprises the sequence of the coding portion of the nucleic acid sequence shown in Figure 1a (SEQ ID NO:1) or the coding portion of the cDNA deposited in NCIMB deposit number 40744 which was deposited at the National Collections of Industrial and Marine Bacteria, 23 St. Machar Drive, Aberdeen AB21RY, Scotland, United Kingdom on June 27, 1995 in accordance with the Budapest Treaty.

A nucleic acid sequence encoding a sodium channel of the present invention may be obtained from a cDNA library derived from mammalian sensory neurons, preferably dorsal root ganglia, trigeminal ganglia or other cranial ganglia, more preferably rat or human dorsal root ganglia. The nucleotide sequence described herein was isolated from a cDNA library derived from rat dorsal root ganglia cells. The nucleic acid sequence coding for the SNS sodium channel has an open reading frame of 5,871 nucleotides encoding a 1,957 amino acid protein. A nucleic acid sequence encoding a sodium channel of the present invention may also be obtained from a mammalian genomic library, preferably a human or rat genomic library. The nucleic acid sequence may be isolated by the subtraction hybridization method described in the examples, by screening with a probe derived from the rat sodium channel sequence, or by other methodologies known in the art such as polymerase chain reaction (PCR) with appropriate primers derived from the rat sodium channel sequence and/or relatively conserved regions of known voltage-gated sodium channels.

30

The nucleic acid sequences of the present invention may be in the form of RNA or in the form of DNA, which DNA includes cDNA, genomic DNA, and synthetic

DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be the coding strand or non-coding (anti-sense) strand. The coding sequence which encodes the rat SNS sodium channel or variant thereof may be identical to the coding sequences set forth herein or that of the deposited clone, or may be a different coding sequence which
5 coding sequence, as a result of the redundancy or degeneracy of the genetic code, encodes the same protein as the sequences set forth herein or the deposited cDNA.

The nucleic acid sequence which encodes the SNS sodium channel may include: only the coding sequence for the full length protein or any variant thereof; the coding sequence for the full length protein or any variant thereof and additional coding
10 sequence such as a leader or secretory sequence or a proprotein sequence; the coding sequence for the full length protein or any variant thereof (and optionally additional coding sequence) and non-coding sequences, such as introns or non-coding sequences 5' and/or 3' of the coding sequence for the full length protein.

The present invention further relates to variants of the hereinabove
15 described nucleic acid sequences which encode fragments, analogs, derivatives or splice variants of the SNS sodium channel. The variant of the SNS sodium channel may be a naturally occurring allelic variant of the SNS sodium channel. As known in the art, an allelic variant is an alternate form of a protein sequence which may have a substitution, deletion or addition of one or more nucleotides, which does not substantially alter the
20 function of the encoded protein. The present invention relates to splice variants of the SNS sodium channel that occur physiologically and which may play a role in changing the activation threshold of the sodium channel.

Variants of the sequence coding for the rat SNS sodium channel have been identified and are listed below:

25 1) a 2573 base pair nucleic acid sequence shown in SEQ ID NO:3. This sequence codes for a 521 amino acid protein that corresponds to amino acids 1437-1957 of Figure 1a (SEQ ID NO:1) and has the same sequence as bases 4512 through 6524 of Figure 1a in the coding portion and 3' untranslated region.

2) a 7052 base pair nucleic acid sequence shown in SEQ ID NO: 5. SEQ
30 ID NO: 6 codes for a 2,132 amino acid protein that contains a 176 amino acid repeat (amino acids 586-760 of SEQ ID NO:6) inserted after amino acid 585 in Figure 1a or SEQ ID NO:2.

A preferred sequence for the rat SNS sodium channel is shown in Figure 1a (SEQ ID NO: 1). However, sequencing variations have been noted. Sequencing has provided

a 6,321 base pair nucleic acid sequence coding for a 1957 amino acid protein that has the same base sequence as bases 1-6321 of Figure 1a or SEQ ID NO:1 with the following changes: bases 1092 G to A, base 1096 C to T, base 2986 G to T, base 3525 C to G and base 3556 G to C.

a 6,527 base pair nucleic acid sequence coding for a 1,957 amino acid protein as shown in SEQ ID NO:7 that has the same base sequence as bases 1-6524 of Figure 1a (SEQ ID NO:1) with an additional 3 bases AAA, at the 3' end, and the following changes: base 299 C to G, base 1092 G to A, base 1096 C to T, base 1964 G to C, base 1965 C to G, base 2472 A to T, base 2986 G to T, base 3019 A to G, base 3158 C to T, base 3525 C to G, base 3556 G to C and base 5893 T to G. The sequence of SEQ ID NO: 7 is also a preferred sequence coding for the rat SNS sodium channel.

a 6524 base pair nucleic acid sequence that has the same sequence as Figure 1a (SEQ ID NO: 1) except for the following base changes: base 1092 G to A (resulting in a change at amino acid 297 of SEQ ID NO: 2 from Val to Ile), base 1096 C to T (resulting in a change at amino acid 298 from Ser to Phe), base 1498 C to A (resulting in a change at amino acid 432 from Ala to Glu), and base 2986 G to T (resulting in a change at amino acid 928 from Ser to Ile).

Sequence variability has been identified in different isolates. One such sequence has been identified that has the sequence of the third sequencing variation shown immediately above except for eight base differences, five of which resulted in an altered amino acid sequence F16-S16, L393-P393, T470-I470, R278-H278, and I1,876-M1,876.

The present invention also relates to nucleic acid probes constructed from the nucleic acid sequences of the invention or portion thereof. Such probes could be utilized to screen a dorsal root ganglia cDNA library to isolate a nucleic acid sequence encoding the sodium channel of the present invention. The nucleic acid probes can include portions of the nucleic acid sequence of the SNS sodium channel or variant thereof useful for hybridizing with mRNA or DNA for use in assays to detect expression of the SNS

sodium channel or localize its presence on a chromosome, such as the *in situ* hybridization assay described herein.

A conservative analogue is a protein sequence which retains substantially the same biological properties of the sodium channel but differs in sequences by one or more conservative amino acid substitutions. For the purposes of this document a conservative amino acid substitution is a substitution whose probability of occurring in nature is greater than ten times the probability of that substitution occurring by chance (as defined by the computational methods described by Dayhoff et al, Atlas of Proteins Sequence and Structure, 1971, page 95-96 and figure 9-10).

A splice variant is a protein product of the same gene, generated by alternative splicing of mRNA, that contains additions or deletions within the coding region (Lewin B. (1995) Genes V Oxford University Press, Oxford, England)

The nucleic acid sequences of the present invention may also have the coding sequence fused in frame to a marker sequence which allows for purification of the protein of the present invention such as a hexa-histidine tag or a hemagglutinin (HA) tag.

The present invention further relates to nucleic acid sequences which hybridize to the hereinabove-described sequences if there is at least 50% and preferably 70% identity between the sequences. The present invention particularly relates to nucleic acid sequences which hybridize under stringent conditions to the hereinabove-described nucleic acid sequences. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences preferably the nucleic acid sequences which hybridize to the hereinabove described nucleic acid sequences encode proteins which retain substantially the same biological function or activity as the SNS sodium channel, however, nucleic acid sequences that have different properties are also within the scope of the present invention. Such sequences, while hybridizing with the above described nucleic acid sequences may encode a protein having different properties, such as sensitivity to tetrodotoxin which property is found in the altered SNS sodium channel protein described herein.

In accordance with another aspect of the invention there is provided purified mammalian sensory neuron sodium channel protein, wherein the sodium channel is insensitive to tetrodotoxin. Preferably the sodium channel of the invention is found in the neurons of the dorsal root ganglia or cranial ganglia, more preferably the neurons of the

dorsal root ganglia. The sodium channel protein may be derived from any mammalian species, preferably the rat or human sodium channel protein. The rat SNS sodium channel protein preferably has the deduced amino acid sequence shown in Figure 1a (SEQ ID NO:2) or SEQ ID NO: 8, or the amino acid sequence encoded by the deposited cDNA.

5 Fragments, analogues, derivatives, and splice variants of the sodium channel specifically located in sensory neurons are also within the scope of the present invention.

The terms "fragment," "derivative" and "analogue" when referring to the DRG sodium channel of the invention refers to a protein which retains substantially the same biological function or activity as such protein. Thus, an analogue includes a
10 proprotein which can be activated by cleavage of the proprotein portion to produce an active mature protein. In addition, the present invention also includes derivatives wherein the biological function or activity of the protein is significantly altered, including derivatives that are sensitive to tetrodotoxin.

The protein of the present invention may be a recombinant protein, a
15 natural protein or a synthetic protein, preferably a recombinant protein.

The fragment, derivative or analog of the SNS sodium channel protein includes, but is not limited to, (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one
20 encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituted group, or (iii) one in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the protein (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature protein, such as a leader or secretory sequence or a sequence which is employed
25 for purification of the mature protein or a proprotein sequence, or (v) one in which one or more amino acids has/have been deleted so that the protein is shorter than the full length protein. Variants of the rat SNS sodium channel are discussed hereinabove and shown in SEQ ID NO:4 and SEQ ID NO:6.

The proteins and nucleic acid sequences of the present invention are
30 preferably provided in an isolated form, and preferably are purified to at least 50% purity, more preferably about 75% purity, most preferably about 90% purity.

The terms "isolated" and/or "purified" mean that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring nucleic acid sequence or protein present in a living animal is not isolated or purified, but the same nucleic acid sequence or DNA or protein, separated
5 from some or all of the coexisting materials in the natural system, is isolated or purified. Such nucleic acid sequence could be part of a vector and/or such nucleic acid sequence or protein could be part of a composition, and still be isolated or purified in that such vector or composition is not part of its natural environment.

The present invention also provides vectors comprising a nucleic acid
10 sequence of the present invention, and host cells transformed or transfected with a nucleic acid of the invention.

The nucleic acid sequences of the present invention may be employed for producing the SNS sodium channel protein or variant thereof by recombinant techniques. Thus, for example, the nucleic acid sequence may be included in any one of a variety of
15 expression vehicles or cloning vehicles, in particular vectors or plasmids for expressing a protein. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences. Examples of suitable vectors include derivatives of SV40; bacterial plasmids; phage DNA; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, pseudorabies and
20 baculovirus. However, any other plasmid or vector may be used as long as it is replicable and viable in the host.

More particularly, the present invention also provides recombinant constructs comprising one or more of the nucleic acid sequences as broadly described above. The constructs comprise an expression vector, such as a plasmid or viral vector,
25 into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises one or more regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of
30 example. Bacterial: pQE70, pQE60, pQE-9 (Qiagen) pBs, phagescript, psiX174, pBluescript SK, pBsKS, pNH8a, pNH16a, pNH18a, pNH461 (Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLneo, pSV2cat.

pOG44, pXT1, pSG (Stratagene), pSVK3, pBPV, pMSG, pSVL (Pharmacia) pcDNA 3.1 (Invitrogen, San Diego, CA), pEE14 (WO 87/04462) and pREP8 (Invitrogen). Preferred vectors include pcDNA 3.1, pEE14 and pREP8. However, any other plasmid or vector may be used as long as it is replicable and viable in the host.

5 As hereinabove indicated, the appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into appropriate restriction endonuclease sites by procedures known in the art. Such procedures and others are deemed to be within the scope of those skilled in the art.

The DNA sequence in the expression vector is operatively linked to an
10 appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. As representative examples of such promoters, there may be mentioned: LTR or SV40 promoter and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression vector may contain a ribosome binding site for translation initiation and transcription terminator. The vector may also include
15 appropriate sequences for amplifying expression.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include LacI, LacZ, T3, T7, gpt, lambda P_R, P_L and trp. Eukaryotic promoters include
20 CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

Depending on the expression system employed in addition, the expression vectors preferably contain a gene to provide a phenotypic trait for selection of transformed
25 host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in *E. coli*.

Transcription of DNA encoding the protein of the present invention by higher eukaryotes can be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp, that act on a
30 promoter to increase its transcription. Examples include the SV40 enhancer on the late side of the replication origin (bp 100 to 270), a cytomegalovirus early promoter enhancer, a polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

Useful expression vectors for bacterial use may be constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but nonlimiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, PKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM1 (Promega Biotec, Madison, Wis., U.S.A.). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

The sodium channel can be expressed in insect cells with the baculovirus expression system which uses baculovirus such as *Autographa Californica* nuclear polyhydrosis virus (AcNPV) to produce large amounts of protein in insect cells such as the Sf9 or 21 clonal lines derived from *Spodoptera frugiperda* cells. See for example O'Reilly et al., (1992) *Baculovirus Expression Vectors: A Laboratory Manual*, Oxford University Press.

Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral

genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

In a further embodiment, the present invention provides host cells capable of expressing a nucleic acid sequence of the invention. The host cell can be, for example, a higher eukaryotic cell, such as a mammalian cell, a lower eukaryotic cell, such as a yeast cell, a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell may be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, electroporation (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, 1986) or any other method known in the art.

Host cells are genetically engineered (transduced, transformed or transfected) with the vectors of this invention which may be, for example, a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the SNS sodium channel genes. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein. As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as *E. coli*, and *Salmonella typhimurium*; Streptomyces; fungal cells, such as yeast; insect cells such as *Drosophila* and *Spodoptera fugiperda* Sf9; animal cells such as CHO, COS or Bowes melanoma Ltk⁺ - and Y1 adrenal carcinoma; plant cells, etc. The selection of an appropriate host is deemed to be within the scope of those skilled in the art based on the teachings herein. Preferred host cells include mammalian cell lines such as CHO-K1, COS-7; Y1 adrenal carcinoma cells. More preferably, the host cells are CHO-K1 cells. Preferred host cells for transient expression of the SNS sodium channel include *Xenopus laevis* oocytes.

The sodium channel may be transiently expressed in *Xenopus laevis* oocytes. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and

expression vectors for use with prokaryotic and eukaryotic hosts are described in Sambrook et al., Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989).

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell, 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, CHO-K1, HeLa, HEK 293, NIH 3T3 and BHK cell lines.

The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the proteins of the invention can be synthetically produced by conventional peptide synthesizers.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well-known to those skilled in the art.

The SNS sodium channel protein is recovered and purified from recombinant cell cultures by methods known in the art, including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, hydroxyapatite chromatography and lectin chromatography. Protein refolding steps may be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

The SNS sodium channel protein of the present invention may be naturally purified products expressed from a high expressing cell line, or a product of chemical synthetic procedures, or produced by recombinant techniques from a prokaryotic or eukaryotic host (for example, by bacterial, yeast, higher plant, insect and mammalian cells in culture).

The present invention also provides antibodies specific for the SNS sodium channel hereinabove defined. The term antibody as used herein includes all immunoglobulins and fragments thereof which contain recognition sites for antigenic

determinants of proteins of the present invention. The antibodies of the present invention may be polyclonal or preferably monoclonal, may be intact antibody molecules or fragments containing the active binding region of the antibody, e.g. Fab or F(ab)₂ and can be produced using techniques well established in the art [see e.g. R.A DeWeger et al; 5 Immunological Rev., 62 p29-45 (1982)].

The proteins, their fragments or other derivatives, or analogs thereof, or cells expressing them can be used as an immunogen to produce antibodies thereto. These antibodies can be, for example, polyclonal or monoclonal antibodies. The present also includes chimeric, single chain and humanized antibodies, as well as Fab fragments, or the 10 product of an Fab expression library. Various procedures known in the art may be used for the production of such antibodies and fragments.

Antibodies generated against the SNS sodium channel can be obtained by direct injection of the polypeptide into an animal or by administering the protein to an animal, preferably a nonhuman. The antibody so obtained will then bind the protein itself. 15 In this manner, even a sequence encoding only a fragment of the protein can be used to generate antibodies binding the whole native protein. Such antibodies can then be used to locate the protein in tissue expressing that polypeptide. For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler and Milstein, 20 1975, Nature 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, Immunology Today 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole, 35 al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss., pp. 77-96).

Techniques described for the production of single chain antibodies (U.S. 25 Pat. No. 4,946,778) can be adapted to produce single chain antibodies to immunogenic polypeptide products of this invention.

The antibodies of the present invention may also be of interest in purifying a protein of the present invention and accordingly there is provided a method of purifying a protein of the present invention as hereinabove defined or any portion thereof or a 30 metabolite or degradation product thereof which method comprises the use of an antibody of the present invention.

The purification method of the present invention may be effected by any convenient technique known in the art for example by providing the antibody on a support and contacting the antibody with a solution containing the protein whereby the antibody binds to the protein of the present invention. The protein may be released from binding
5 with the antibody by known methods for example by changing the ionic strength of the solution in contact with the complex of the protein/antibody.

The present invention also provides methods of identifying modulators of the sodium channel which is specifically located in sensory neurons comprising contacting a test compound with the sodium channel and detecting the activity of the sodium channel.
10 Preferably, the methods of identifying modulators or screening assays employ transformed host cells that express the sodium channel. Typically, such assays will detect changes in the activity of the sodium channel due to the test compound, thus identifying modulators of the sodium channel. Modulators of the sodium channel are useful in modulating the sensation of pain. Blockers of the sodium channel will prevent the transmission of
15 impulses along sensory neurons and thereby be useful in the treatment of acute, chronic or neuropathic pain.

The sodium channel can be used in a patch clamp or other type of assay, such as the assays disclosed herein in the examples, to identify small molecules, antibodies, peptides, proteins, or other types of compounds that inhibit, block, or otherwise interact
20 with the sodium channel. Such modulators identified by the screening assays can then be used for treatment of pain in mammals.

For example, host cells expressing the SNS sodium channel can be employed in ion flux assays such as $^{22}\text{Na}^+$ ion flux and ^{14}C guanidinium ion assays, as described in the examples and in the art, as well as the SFBI fluorescent sodium indicator
25 assays as described in Levi et al., (1994) J. Cardiovascular Electrophysiology 5:241-257. Host cells expressing the SNS sodium channel can also be employed in binding assays such as the 3H-batrachotoxin binding assay described in Sheldon et al., (1986) Molecular Pharmacology 30:617-623; the 3H-saxitoxin assay as described in Rogart et al (1983) Proc. Natl. Acad. Sci. USA 80:1106-1110; and the scorpion toxin assay described in West et al.,
30 (1992) Neuron 8:59-70. Additionally, the host cells expressing the SNS sodium channel can be used in electrophysiological assays using patch clamp or two electrode techniques. In general, a test compound is added to the assay and its effect on sodium flux is

determined or the test compound's ability to competitively bind to the sodium channel is assessed. Test compounds having the desired effect on the SNS sodium channel are then selected. Modulators so selected can then be used for treating pain as described above.

Complementary strands of the nucleotide sequences as hereinabove defined
5 can be used in gene therapy, such as disclosed in U.S. Patent 5,399,346. For example, the cDNA sequence or fragments thereof could be used in gene therapy strategies to down regulate the sodium channel. Antisense technology can be used to control gene expression through triple-helix formation or antisense DNA or RNA, both of which methods are based on binding of a nucleic acid sequence to DNA or RNA. For example, the 5' coding portion
10 of the nucleic acid sequence that encodes the sodium channel is used to design an antisense RNA oligonucleotide of from about 10 to about 40 base pairs in length. A DNA oligonucleotide is designed to be complimentary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al, Science 241:456 (1988); and Deruau et al., Science 251:1360 (1991)), thereby preventing
15 transcription and the product of the sodium channel. The antisense RNA oligonucleotide hybridizes to the mRNA in vivo and blocks translation of the mRNA into the sodium channel. Antisense oligonucleotides or an antisense construct driven by a strong constitutive promoter expressed in the target sensory neurons would be delivered either peripherally or to the spinal cord.

20 The regulatory regions controlling expression of the sodium channel gene could be used in gene therapy to control expression of a therapeutic construct in cells expressing the sodium channel.

Such regions would be isolated by using the cDNA as a probe to identify genomic clones carrying the gene and also flanking sequence e.g. cosmids. Fragments of
25 the cosmids containing intron or flanking sequence would be used in a reporter gene assay in e.g. DRG cultures or transgenic animals and genomic fragments carrying e.g. promoter, enhancer or LCR activity identified.

The invention will now be further described with reference to the following examples:

30 **Example 1 - Derivation of the sequence of a rat dorsal root ganglia (DRG) sodium channel cDNA by subtraction hybridisation methodology**

1.1 cDNA synthesis from DRG-derived poly-A+ RNA

Dorsal root ganglia (DRG) from all spinal levels of neonatal Sprague-Dawley male and female rats were frozen in liquid nitrogen. RNA is extracted
5 using guanidine isothiocyanate and phenol/chloroform extraction (Chomczynski and Sacchi 1987 Anal Biochem 162,156-159).

Total RNA isolation - the nerve tissue is homogenised using a Polytron homogeniser in 1ml extraction buffer (23.6g guanidinium isothiocyanate, 5ml of 250 mM sodium citrate (pH 7.0) made up to 50ml with distilled water. To this is added 2.5ml 10%
10 sarcosyl and 0.36ml β -mercaptoethanol). 0.1ml of 2M sodium acetate (pH 4.0) is added followed by 1 ml phenol. After mixing, 0.2ml chloroform is added and this is shaken vigorously and placed on ice for 5 minutes. This is then centrifuged at 12,000 revolutions per minute (rpm) for 30 minutes at 4°C. The aqueous phase is transferred to a fresh tube, 1ml of isopropanol is added and this is left at -20°C for an hour followed by centrifuging at
15 12000 rpm for 30 minutes at 4°C. The pellet is dissolved in 0.1ml extraction buffer and is again extracted with isopropanol. The resulting pellet is washed with 70% ethanol and is resuspended in diethyl pyrocarbonate (DEPC)-treated water. 0.3M sodium acetate (pH5.2) and 2 volumes of ethanol are added and the mixture is placed at -20°C for 1 hour. The RNA is precipitated, washed again with 70% ethanol and resuspended in DEPC-treated
20 water. The optical density is measured at 260 nanometres (nm) to calculate the yield of total RNA. Poly A+ RNA is isolated from the total RNA by oligo-dT cellulose chromatography (Aviv and Leder 1972 Proc Natl Acad Sci 69.1408-1411). The following procedures are carried out at 4°C as far as is possible. Oligo-dT cellulose (Sigma) is prepared by treatment with 0.1M sodium hydroxide for 5 minutes. The oligo-dT resin is
25 poured into a column and is neutralised by washing with neutralising buffer (0.5 M potassium chloride, 0.01M Tris (Trizma base - Sigma - Tris(hydroxymethyl)aminomethane) (pH 7.5). The RNA solution is adjusted to 0.5M potassium chloride, 0.01M Tris (pH7.5) and is applied to the top of the column. The first column eluate is re-applied to the column to ensure sticking of the mRNA to the oligo-dT
30 in the column. The column is then washed with 70ml of neutralising buffer and the polyA+ RNA is eluted with 6ml 0.01M Tris (pH7.5) and 1ml fractions are collected. The poly A+ RNA is usually in fractions 2 to 5 and this is checked by measuring the optical density at

260nm. These fractions are pooled and ethanol precipitated overnight at -70°C, washed in 70% ethanol and then redissolved in deionised water at a concentration of 1mg/ml.

First strand cDNA was generated using 0.5mg DRG poly A+ mRNA, oligo-dT/Not-I primer adapters and SuperScript reverse transcriptase (Gibco-BRL) using methodology as described in example 2. One half of the cDNA was labelled by including 2 MBq ³²P dCTP (Amersham) in the reverse transcriptase reaction. Labelled cDNA is separated from unincorporated nucleotides on Nick columns (Sephadex G50 - Pharmacia).

1.2 Enrichment of DRG-specific cDNA using subtraction hybridisation.

10

Poly A+ RNA from various tissues (10µg) is incubated with 10µg photoactivatable biotin (Clontech) in a total volume of 15µl and irradiated at 4°C for 30 minutes with a 250 watt sunlamp. The photobiotin is removed by extraction with butanol, and the cDNA co-precipitated with the biotinylated RNA without carrier RNA (Sive and St. John 1988 Nuc Ac Res 16,10937).

Hybridisation is carried out at 58°C for 40 hours in 20% formamide, 50mM 3-(N-morpholino)propanesulphonic acid (MOPS) (pH 7.6), 0.2% sodium dodecyl sulphate (SDS), 0.5M sodium chloride, 5mM ethylenediaminetetraacetate (EDTA - Sigma). The total reaction volume is 5µl and the reaction is carried out under mineral oil, after an initial denaturation step of 2 minutes at 95°C. 100µl 50mM MOPS (pH 7.4), 0.5M sodium chloride, 5mM EDTA containing 20 units of streptavidin (BRL) is then added to the reaction mixture at room temperature. and the aqueous phase retained after two phenol /chloroform extraction steps. After sequential hybridisation of the cDNA from Example 1.1 with biotinylated mRNA from liver and kidney, followed by cortex and cerebellum, a 80-fold concentration of DRG-specific transcripts is achieved.

One third of the 1-2 ng of residual cDNA is then G-tailed with terminal deoxynucleotide transferase at 37°C for 30 minutes. The polymerase chain reaction is used to amplify the cDNA using an oligo-dT-Not-I primer adapter and oligo-dC primers starting with the sequence AATTCCGA(C)₁₀. Amplification is carried out using 2 cycles of 95°C for 1min, 45°C for 1 min, 72°C for 5min, followed by 2 cycles of 95°C for 1 minute, 58°C for 1 minute and 72°C for 5 minutes. The resulting products are then separated on a

2% Nu-sieve agarose gel, and material running at a size of greater than 0.5 kilobase pairs (kb) is eluted and further amplified with 6 cycles of 95°C for 1 minute, 58°C for 1 minute and 72°C for 5 minutes. This material is further separated on a 2% Nu-sieve agarose gel, and the material running from 6kb on the gel is eluted and further amplified using the same
5 PCR conditions for 27 cycles. The amplified DNA derived from this high molecular weight region is then further fractionated on a 2 % Nu-Sieve gel, and cDNA from 0.5 to 1.5kb, and from 1.5 to 5kb pooled.

1.3. Library Construction

10 10µg of the bacteriophage vector lambda-zap II (Stratagene) is restriction digested with NotI and EcoRI in high salt buffer overnight at 37°C followed by dephosphorylation using 1 unit of calf intestinal phosphatase (Promega) for 30 minutes at 37°C in 10mM Tris.HCl (pH9.5), 1mM spermidine, 0.1mM EDTA. DRG cDNA is digested with Klenow enzyme in the presence of dGTP and dCTP to construct an EcoRI
15 site from the oligo-dC primer (see above) at the 5' end of the cDNA, and cut with NotI for directional cloning. The cDNA is ligated into the cloning vector bacteriophage lambda-zap II for 16 hours at 12°C. Recombinant phage DNA is then packaged into infective phage using Gigapack gold (Stratagene) and protocols specified by the suppliers. 0.1% of the packaged DNA is used to infect E.coli BB4 cells which are plated out to
20 calculate the number of independent clones generated.

1.4 Differential Screening

The library is plated at a low density (10^3 clones/ 12 x 12 cm² dish) and
25 screened using three sets of ³²P-labelled cDNA probes and multiple filter lifts. Replica filters are made by laying them onto the plated library plates, briefly drying them and then laying onto fresh agar plates to increase the quantity of phage and the subsequent hybridisation signals of lifts taken from them. The probes are derived from: a) cortex and cerebellum poly (A)+ RNA, b) DRG poly (A)+ RNA, and c) subtracted cDNA from
30 DRG. The two mRNA probes are labelled with ³²P dCTP using a reaction mixture containing 2-5µg RNA, 50µl 5 x RT buffer, 25 µl 0.1M dithiothreitol (DTT), 12.5µl

10mM dATP, dGTP, dCTP, 30pM oligo-dT, 75 μ l 32 P-dCTP (30MBq; Amersham), 25 μ l 100 μ M dCTP, 2 μ l RNasin (2units/ μ l) and 2 μ l SuperScript reverse transcriptase (GibcoBRL) in a final volume of 250 μ l. The reaction is incubated at 39°C for 60 minutes, and the RNA subsequently destroyed by adding 250 μ l water, 55 μ l 1M NaOH, and

5 incubating at 70°C for 20 minutes. The reaction mixture is neutralised with acidified Tris base (pH 2.0) and precipitated with carrier tRNA (Boehringer) with isopropanol. The subtracted and amplified double-stranded DRG cDNA is random-prime labelled with 32 P dATP (Gibco multiprime kit). Replica filters are then prehybridised for 4 hours at 68°C in hybridisation buffer. Hybridisation was carried out for 20 hours at 68°C in 4x SSC

10 (20xSSC consists of 175.3g of sodium chloride and 88.2g of sodium citrate in 800ml of distilled water. The pH is adjusted to 7.0 with 10N sodium hydroxide and this is made to 1 litre with distilled water), 5x Denhardt's solution containing 150 μ g/ml salmon sperm DNA, 20 μ g/ml poly-U, 20 μ g/ml poly-C, 0.5% SDS (Sigma), 5mM EDTA. The filters are briefly washed in 2 x SSC at room temperature, then twice with 2 x SSC with 0.5% SDS at 68°C

15 for 15 minutes, followed by a 20 minute wash in 0.5% SDS, 0.2 x SSC at 68°C. The filters are autoradiographed for up to 1 week on Kodak X-omat film. Plaques that hybridise with DRG probes but not cortex and cerebellum probes are picked, phage DNA prepared and the cloned inserts released for subcloning into pBluescript (Stratagene).

The positive plaques are picked by lining up the autoradiogram with the

20 plate using orientation marks and taking a plug from the plate corresponding to the positive hybridisation signal. The phage is eluted from the plug in 0.5ml phage dilution buffer (10mM Tris chloride (pH7.5) 10mM magnesium sulphate) and the phage re-infected into E.coli BB4 and replated at a density of 200 to 1000 plaques/150mm plate as a secondary purification step to ensure purity of the clones. The positive secondaries are then picked as

25 described previously. In order to sub-clone the insert DNA from the positive recombinant phage, they need to be amplified. This is accomplished by plate lysis where the phage totally lyse the E.coli BB4. 0.2ml of phage suspension is mixed with 0.1ml of an overnight culture of E.coli. This is added to 2.5ml of top agar (16g bacto-tryptone 10g bacto-yeast extract, 5g sodium chloride, 7g bacto-agar in 900mls distilled water) and plated onto 9cm²

30 agar plates. These are incubated overnight at 37°C. 5ml of phage dilution buffer is then added to the plates and is incubated overnight at 4°C or for 4 hours with gentle scraping at

room temperature. The phage-containing buffer is then recovered, 0.1ml chloroform is added and this phage stock is titrated as above and stored at 4°C. Phage DNA is prepared by first infecting 10¹⁰ E.coli B44 with 10⁹ plaque forming units (pfus) of phage in 3ml of phage dilution buffer and shaking at 37°C for 20 minutes. The infected bacteria are added
5 to 400ml of L broth (1.6% bactotryptone, 0.5% (w/v) Bacto yeast extract, 0.5% (w/v) magnesium sulphate) with vigorous shaking at 37°C for 9 hours. When lysis has occurred, 10ml of chloroform is added and shaking is continued for a further 30 minutes. The culture is then cooled to room temperature and pancreatic RNAase and DNAase are added to 1ug/ml for 40 minutes. Sodium chloride is then added to 1M and is dissolved by swirling
10 on ice. After centrifuging at 8000rpm for 10 minutes the supernatant is recovered. Polyethylene glycol (PEG 6000) is added to 10% w/v and is dissolved by stirring whilst on ice for 2 hours. After centrifuging for 8000rpm for 10 minutes at 4°C the pellet is resuspended in 8ml of phage dilution buffer. This is extracted with an equal volume of phenol/chloroform followed by purification on a caesium chloride gradient (0.675g/ml
15 caesium chloride - 24 hours at 38000 rpm at 4°C). The opaque phage band is removed from the centrifugation tube and dialysed against 10mM sodium chloride, 50mM Tris (pH8.0), 10mM magnesium chloride for 2 hours. EDTA is then added to 20mM, proteinase K to 50µg/ml and SDS to 0.5% and is incubated at 65°C for 1 hour. After dialysis overnight against TE pure phage DNA results. The cloned insert is digested from the
20 purified phage DNA using restriction enzymes as previously described. Each phage insert is then ligated into a plasmid vector e.g. pBluescript - Clontech using a ligation reaction as previously described.

Clone characterisation.

25

The plasmids are cross hybridised with each other. Unique clones are further analysed by Northern blotting and sequencing. The clone/s showing transcript sizes and sequence comparable with sodium channels are then used as hybridisation probes to screen a neonatal rat DRG oligo dT-primed full length cDNA library to derive full length cDNA
30 clones using methodology as described above and in example 2. Biological activity of the rat DRG sodium channel is confirmed as in examples 4 and 7 below.

Example 2 - Homology cloning of the human cDNA homologous to the rat DRG sodium channel cDNA (SNS-B).

2.1. Isolation of human ganglia total RNA

5

The starting material for the derivation of the human cDNA homologue of the rat DRG sodium channel cDNA is isolated human dorsal root ganglia or trigeminal ganglia or other cranial ganglia from post-mortem human material or foetuses. Total ribonucleic acid (RNA) is isolated from the human neural tissue by extraction in
10 guanidinium isothiocyanate (Chomczynski and Sacchi 1987 Anal Biochem 162,156-159) as described in example 1.

2.2 Determination of the transcript size of the human homologue of the rat DRG sodium channel cDNA (SNS-B).

15

Human dorsal root ganglia total RNA is electrophoretically separated in a 1% (w/v) agarose gel containing a suitable denaturing agent e.g. formaldehyde (Lehrach et al 1977 Biochemistry 16,4743; Goldberg 1980 Proc Natl Acad Sci 77,5794; Seed 1982 in Genetic engineering: principles and methods (ed JK Setlow and A Hollaender) vol 4 p91
20 Plenum Publishing New York) or glyoxal/DMSO (McMaster GK and Carmichael GG 1977 Proc Natl Acad Sci 74,4835), followed by transfer of the RNA to a suitable membrane (e.g. nitrocellulose). The immobilised RNA is then hybridised to radioactive (or other suitable detection label) probes consisting of portions of the rat sodium channel cDNA sequence (see below). After washing of the membrane to remove non-hybridised
25 probe, the hybridised probe is visualised using a suitable detection system (e.g. autoradiography for ³²P labelled probes) thus revealing the size of the human homologous mRNA molecule. Specifically, 20-30 µg total RNA from neonatal rat tissues are separated on 1.2% agarose -formaldehyde gels, and capillary blotted onto Hybond-N (Amersham) (Ninkina et al. 1993 Nuc Ac Res 21,3175-3182). The amounts of RNA on the blot are
30 roughly equivalent, as judged by ethidium bromide staining of ribosomal RNA or by hybridisation with the ubiquitously expressed L-27 ribosomal protein transcripts (Le Beau et al. 1991 Nuc Ac Res 19,1337). Each Northern blot contains human DRG, cortex, cerebellum, liver kidney, spleen and heart RNA. Probes (50ng) are labelled with ³²P-dATP

(Amersham) by random priming. Filters are prehybridised in 50% formaldehyde 5 x SSC containing 0.5% SDS, 5 x Denhardts solution (50x Denhardts contains 5g of Ficoll (Type 400, Pharmacia), 5g of polyvinylpyrrolidone, 5g of bovine serum albumin (Fraction V, Sigma) and water to 500ml), 100 µg/ml boiled salmon sperm DNA, 10 µg/ml poly-U and
5 10 µg/ml poly-C at 45°C for 6 hours. After 36 hours hybridisation in the same conditions, the filters are briefly washed in 2 x SSC at room temperature, then twice with 2 x SSC with 0.5% SDS at 68°C for 15 minutes, followed by a 20 minute wash in 0.5% SDS, 0.2 x SSC at 68°C. The filters are autoradiographed for up to 1 week on Kodak X-omat film. The
10 transcript size is calculated from the signal from the gel in comparison with gel molecular weight standard markers.

2.3 Production of a human DRG cDNA library

In order to produce a representative cDNA library from the human dorsal
15 root ganglia messenger RNA (poly A+ mRNA) is first isolated from the total RNA pool using oligo-dT cellulose chromatography (Aviv and Leder 1972 Proc Natl Acad Sci 69,1408-1411) using methodology described in example 1. Synthesis of the first strand of cDNA from the polyA+ RNA uses the enzyme RNA-dependent DNA polymerase (reverse transcriptase) to catalyse the reaction. The most commonly used method of second strand
20 cDNA synthesis uses the product of first strand synthesis, a cDNA:mRNA hybrid, as a template for priming the second strand synthesis. (Gubler and Hoffman 1983 Gene 25,263)).

2.3.1. First strand cDNA synthesis

25

20µg of human DRG polyA+ RNA is pre-treated to destroy secondary structure which may inhibit first strand cDNA synthesis. 20µg of polyA+ RNA, 1µl 1M Tris (pH7.5) are made up to a volume of 100µl with distilled water. This is incubated at 90°C for 2 minutes followed by cooling on ice. 4.8 µl of 100 mM methyl mercury is then
30 added for 10 minutes at room temperature. 10µl of 0.7M β-mercaptoethanol and 100 units of human placental RNAase inhibitor are then added for 5 minutes at room temperature.

The first strand synthesis reaction consists of 8µl 20mM dATP, 5µl 20mM dCTP, 8µl 20mM dGTP 8µl 20mM dTTP, 10µl 1mg/ml oligo-dT (12-18), 20µl 1M Tris (pH 8.3) (at 45°C), 8µl 3M potassium chloride, 3.3µl 0.5M magnesium chloride, 3µl $a^{32}P$ dCTP, 100 units Superscript II reverse transcriptase (GibcoBRL) made up to 200µl with distilled water. This reaction mixture is incubated at 45°C for 45 minutes after which another 50 units of Superscript reverse transcriptase is added and incubated for a further 30 minutes at 45°C. EDTA is then added to 10mM to terminate the reaction and a phenol/chloroform extraction is carried out. The DNA is then precipitated using ammonium acetate (freezing in dry ice/ethanol before centrifuging), washed with 70% ethanol and resuspended in 50ml distilled water. The size of the single stranded DNA is assessed by electrophoretically separating it out on an agarose gel (1% w/v) and autoradiographing the result against markers.

2.3.2 Second strand synthesis

The second strand synthesis reaction mixture consists of 0.5µg human DRG single stranded DNA, 2µl 1M Tris (pH7.5), 1µl 0.5M magnesium chloride, 3.33µl 3M potassium chloride, 2µl 0.5M ammonium sulphate, 1.5µl 10mM βnicotinamide adenine dinucleotide (NAD), 4µl of each of the 1mM dNTPs, 5µl 1mg/ml bovine serum albumin (BSA), 1 unit RNAase-H, 25 units Klenow polymerase all made up to 100µl with distilled water. This is incubated at 12°C for 1 hour and then at 20°C for 1 hour. The reaction is stopped by addition of EDTA to 20mM followed by a phenol/chloroform extraction. The DNA is ethanol precipitated (-70°C overnight) and is then washed with 70% ethanol followed by resuspension in 20µl distilled water. Size is checked by gel electrophoresis and autoradiography.

2.3.3 Double stranded cDNA end repair

In order to add linkers to the end of the cDNA molecules for subsequent cloning, the ends must first be repaired. The human DRG cDNA is treated with 500 units/ml of S1 nuclease in 0.25M sodium chloride, 1mM zinc sulphate, 50mM sodium

acetate (pH4.5). Incubation is at 30°C for 40 minutes followed by neutralisation with Tris (pH 8.0) to 0.2M. The DNA is again ethanol precipitated, washed in 70% ethanol and resuspended in 20ul distilled water. The size is again checked to ensure that S1 nuclease digestion has not radically reduced the average DNA fragment size. The repair reaction
5 consists of 19µl cDNA, 3µl 10xT4 polymerase buffer (0.33M Tris acetate (pH7.9), 0.66M potassium acetate, 0.1M magnesium acetate, 1mg/ml BSA and 5mM DTT), 2µl of each dNTP at 2mM, 2µl T4 polymerase and 4µl distilled water. This is incubated at 37°C for 30 minutes followed by addition of 1µl Klenow polymerase for 1 hour at room temperature. The DNA is then ethanol precipitated, washed in 70% ethanol and resuspended in 5µl
10 distilled water. In order to protect naturally occurring restriction sites within the cDNA from being cleaved, the cDNA is treated with a methylase before the addition of linkers. The reaction mixture consists of 5µl human DRG double stranded DNA, 1µl S-adenosylmethionine, 2µl 1mg/ml BSA, 2µl 5x methylase buffer (0.5M Tris (pH8.0), 5mM EDTA), 0.2µl EcoRI methylase (NEB). This is incubated at 37°C for 20 minutes followed
15 by phenol extraction, ethanol precipitation washing with 70% ethanol and resuspension in 20µl distilled water.

2.3.4. Addition of linkers to cDNA

20 EcoRI linkers are ligated to the cDNA molecules to facilitate cloning into lambda vectors. The ligation reaction mixture consists of 1µl 10x ligation buffer (0.5M Tris chloride (pH7.5), 0.1M magnesium chloride and 0.05M DTT), 1µl 10mM ATP, 100ng cDNA, 5µg EcoRI linkers, 1 unit T4 DNA ligase, distilled water to 10µl. The reaction is incubated at 37°C for 1 hour, followed by addition of 6 more units of T4 ligase and a
25 further incubation overnight at 15°C. The ligated samples are ethanol precipitated, washed in 70% ethanol and resuspended in 10µl distilled water. The cDNA is then digested with EcoRI to cleave any linker concatamers formed in the ligation process. This restriction digestion reaction contains 10µl cDNA, 2µl high salt buffer (10mM magnesium chloride, 50mM Tris chloride (pH7.5), 1mM DTT, 100mM sodium chloride), 2µl EcoRI (10 units/µl
30 - NEB) and distilled water to 20µl. The digestion is carried out for 3 hours. The ligation

and digestion steps are monitored using gel electrophoresis to monitor the size of the products.

5 2.3.5 Size fractionation of cDNA

In order to assure that the library is not swamped with short cDNA molecules and to remove linker molecules a column purification is carried out. A 1ml Sepharose 4B column is made in a 1 ml plastic pipette plugged with a small piece of glass
10 wool. This is equilibrated with 0.1M sodium chloride in TE. The cDNA is loaded onto the column and 1 drop fractions are collected. 2µl aliquots of each fraction are analysed by gel electrophoresis and autoradiography to determine the sizes of the cDNA in each fraction. Fractions containing cDNA of about 800 base pairs and above are pooled and purified by ethanol precipitation and resuspending in 10µl distilled water.

15

2.3.6 Cloning of cDNA into bacteriophage vector

Bacteriophage vectors designed for the cloning and propagation of cDNA are provided ready-digested with EcoRI and with phosphatased ends from commercial
20 sources (e.g. lambda gt10 from Stratagene). The prepared subtracted cDNA is ligated into lambda gt10 using a ligation reaction consisting of ligase buffer and T4 DNA ligase (New England Biolabs) as described elsewhere in this document.

2.4 Labelling of cDNA fragments (probes) for library screening

25

The 3' untranslated region of the rat DRG sodium channel cDNA clone (SNS-B) is subcloned using appropriate restriction enzymes into a plasmid vector e.g. pBluescript - Stratagene. The cDNA insert which is to form the labelled probe is released from the vector via digestion with appropriate restriction enzymes and the insert is
30 separated from the vector via electrophoresis in a 1% (w/v) agarose gel. After removal of the separated insert from the agarose gel and purification it is labelled by standard

techniques such as random priming and polymerisation (Feinberg and Vogelstein 1983 Anal Biochem 132,6) or nick translation (Rigby et al 1977 J Mol Biol 113,237) with ^{32}P or DIG-labelled nucleotides. Alternatively, if the probe cDNA insert is cloned into a vector containing strong bacteriophage promoters to which DNA-dependant RNA polymerases
5 bind (SP6, T3 or T7 polymerases), synthetic cRNA is produced by in vitro transcription which incorporates ^{32}P or digoxigenin nucleotides. Other regions of the rat DRG sodium channel cDNA can also be used as probes in a similar fashion for cDNA library screening or Northern blot analysis. Specifically, a probe is made using a kit such as the Pharmacia oligo labelling kit. This will radioactively label the rat DRG sodium channel cDNA
10 fragment. 50ng of denatured DNA (place in boiling waterbath for 5 minutes), 3 μl of ^{32}P dCTP (Amersham) and 10 μl reagent mix is made up to 49 μl with distilled water. 1 μl of Klenow fragment is added and the mixture is incubated at 37°C for one hour. To remove unincorporated nucleotides, the reaction mixture is applied to a Nick column (Sephadex G50 - Pharmacia) followed by 400 μl of TE (10mM Tris chloride (pH7.4) 1mM EDTA
15 (pH8.0)). Another 400 μl of TE is added and the eluate is collected. This contains the labelled DNA to be used as a hybridisation probe.

2.5 cDNA library screening

20 In order to detect recombinants containing human homologues of the rat DRG sodium channel the human DRG cDNA library is screened using moderate stringency hybridisation washes (50-60°C, 5 x SSC, 30 minutes), using radiolabelled or other labelled DNA or cRNA probes derived from the 3' untranslated region as described above. Libraries are screened using standard methodologies involving the production of
25 nitrocellulose or nylon membrane replicas of DNA from recombinant plaques formed on agar plates (Benton et al 1977 Science 196:180). These are then hybridised to single stranded nucleic acid probes (see above). Moderate stringency washes are carried out (see wash conditions for Northern analysis in section 2.2). Plaques which are positive on duplicate filters (i.e. not artefacts or background) are then purified by one or more rounds
30 of replating after dilution to separate the colonies and further hybridisation screening. Resulting positive plaques are purified. DNA is extracted and the insert sizes of these

AUG 24 2000

clones is examined. The clones are cross-hybridised to each other using standard techniques (Sambrook et al 1989 Molecular Cloning Second Edition Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York) and distinct positive clones identified. Detailed protocols for cDNA library screening are given in example 1.

5

2.6 Derivation of a full-length clone of the human homologue of the rat DRG sodium channel cDNA

Overlapping positive clones from above are identified by cross-hybridisation. They are then restriction mapped to identify their common portions and restriction fragments representing the separate portions from the overlapping clones are ligated together using standard cloning techniques (Sambrook et al 1989 Molecular Cloning Second Edition Cold Spring Harbor Laboratory Press). For example, the most 5' fragment will contain any 5' untranslated sequence, the start codon ATG and 5' coding sequence. The most 3' clone will contain the most 3' coding sequence, a stop codon and any 3' untranslated sequence, a poly A consensus sequence and possibly a poly A run. Thus a recombinant molecule is generated which contains the full cDNA sequence of the human homologue of the rat DRG sodium channel cDNA. If overlapping clones do not produce sufficient fragments to assemble a full length cDNA clone, the full length oligo dT-primed human DRG library is re-screened to isolate a full length clone. Alternatively, a full length clone is derived directly from the library screening.

2.7 Characterisation of the human homologue full-length clone

The cDNA sequence from the full-length clone is used as a probe in Northern blot analysis to detect the messenger RNA size in human tissue for comparison with the rat messenger RNA size (see sections 1.1 and 2.2 for methodology).

Confirmation of biological activity of the cloned cDNA is carried out via in vitro translation of the human sodium channel mRNA and its expression in *Xenopus* oocytes in an analogous manner to that for the rat DRG-specific TTXi resistant sodium channel as described in examples 4 and 7.

cDNA sequences which are shown to have activity as defined above are completely sequenced using dideoxy-mediated chain termination sequencing protocols (Sanger et al 1977 Proc Natl Acad Sci 74,5463).

5 **Example 3 - Polymerase chain reaction (PCR) approaches to clone the human DRG sodium channels using DNA sequence derived from the rat DRG sodium channel cDNA clone**

10 Total RNA and poly A+ RNA is isolated from human dorsal root ganglia or trigeminal ganglia or other cranial ganglia from post-mortem human material or foetuses as described in example 2 above.

Random primers are hybridised to the RNA followed by polymerisation with MMLV reverse transcriptase to generate single stranded cDNA from the extracted human RNA.

15 Using degenerate PCR primers derived from relatively conserved regions of the known voltage-gated sodium channels (Figure 2), amplify the cDNA using the polymerase chain reaction (Saiki et al 1985 Science 230,1350). It is appreciated by those skilled in the art that there are many variables which can be manipulated in a PCR reaction to derive the homologous sequences required. These include but are not limited to varying
20 cycle and step temperatures, cycle and step times, number of cycles, thermostable polymerase, Mg²⁺ concentration. It is also appreciated that greater specificity can be gained by a second round of amplification utilising one or more nested primers derived from further conserved sequence from the sodium channels.

Specifically, the above can be accomplished in the following manner. The
25 first strand cDNA reaction consists of 1µg of total RNA made up to 13µl with DEPC-treated water and 1µl of 0.5µg/µl oligo(dT). This is heated to 70°C for 10 minutes and then incubated on ice for 1 minute. The following is then added: 2µl of 10x synthesis buffer (200mM Tris chloride, 500mM potassium chloride, 25mM magnesium chloride, 1µg/ml BSA), 2µl of 0.1M DTT, 1µl of 200U/µl Superscript Reverse Transcriptase (Gibco
30 BRL). This is incubated at room temperature for 10 minutes then at 42°C for 50 minutes. The reaction is then terminated by incubating for 15 minutes at 70°C. 1µl of E.coli RNase H (2U/µl) is added to the tube which is then incubated for 20 minutes at 37°C.

The PCR reaction is set up in a 0.5ml thin-walled Eppendorf tube. The following reagents are added: 10µl 10x PCR buffer, 1µl cDNA, 16µl dNTP's (25µl of 100µM dATP, dCTP, dGTP and dTTP into 900µl sterile distilled water), 7µl of 25mM magnesium chloride, 1µl of Taq DNA polymerase (Amplitaq Perkin-Elmer) plus sterile distilled water to 94µl.

To each reaction tube a wax PCR bead is added (Perkin-Elmer) and the tube placed in a 70°C hot block for 1 minute. The tubes are allowed to cool until the wax sets and 3µl of each primer (33pM/µl) are added above the wax. The tubes are placed in a thermal cycler (Perkin-Elmer) and the following 3-step program used after an initial 94°C for 5 minutes; 92°C for 2 minutes, 55°C for 2 minutes, 72°C for 2 minutes for 35 cycles. A final polymerisation step is added at 72°C for 10 minutes. The reaction products are then run on a 1% agarose gel to assess the size of the products. In addition, control reactions are performed alongside the samples. These should be: 1) all components without cDNA (negative control) and 2) all reaction components with primers for constitutively expressed product e.g. α -actin or HPRT.

The products of the PCR reactions are examined on 0.8%-1.2% (w/v) agarose gels. Bands on the gel (visualised by staining with ethidium bromide and viewing under UV light) representing amplification products of the approximate predicted size were then cut from the gel and the DNA purified. Further bands of interest are also identified by Southern blot analysis of the amplification products and probing of the resulting filters with labelled primers from further conserved regions e.g. those used for secondary amplification.

The resulting DNA is ligated into suitable vectors such as, but not limited to, pCR II (Invitrogen) or pGemT. Clones are then sequenced to identify those containing sequence with similarity to the rat DRG sodium channel sequence (SNS-B).

Clone analysis

Candidate clones from above are used to screen a human cDNA DRG library constructed using methods described in example 2. If a full length clone is not identified, positive overlapping clones which code for the full length human cDNA

homologue are identified and a full length clone is then assembled as described in example 1. Biological activity is then confirmed as described in examples 4 and 7.

Example 4 - In vitro translation of rat and human DRG sodium channel in *Xenopus laevis* oocytes

5 In order to demonstrate the biological activity of the protein coded for by the rat DRG sodium channel cDNA sequence (SNS-B) and its human homologue the complete double-stranded cDNA coding sequences are ligated into in vitro transcription vectors
10 (including but not limited to the pGEM series, Promega) using one or more of the available restriction enzyme sites such that the cDNAs are inserted in the correct orientation. The constructs are then used to transform bacteria and constructs with the correct sequence in the correct orientation are identified via diagnostic restriction enzyme analysis and dideoxy-mediated chain termination DNA sequencing (Sanger et al 1977 Proc Natl Acad
15 Sci 74,5463).

These constructs are then linearised at a restriction site downstream of the coding sequence and the linearised and purified plasmids are then utilised as a template for in vitro transcription. Sufficient quantities of synthetic mRNA are produced via in vitro transcription of the cloned DNA using a DNA-dependent RNA polymerase from a
20 bacteriophage that recognises a bacteriophage promoter found in the cloning vector. Examples of such polymerases include (but are not limited to) T3, T7 and SP6 RNA polymerase.

A variation on the above method is the synthesis of mRNA containing a 5' terminal cap structure (7-methylguanosine) to increase its stability and enhance its
25 translation efficiency (Nielson and Shapiro 1986 Nuc Ac Res 14,5936). This is accomplished by the addition of 7-methylguanosine to the reaction mixture used for synthetic mRNA synthesis. The cap structure is incorporated into the 5' end of the transcripts as polymerisation occurs. Kits are available to facilitate this process e.g. mCAP RNA Capping Kit - Stratagene).

30 The synthetic RNA produced from the in vitro transcription is isolated and purified. It is then translated via microinjection into *Xenopus laevis* oocytes. 50nl of 1mg/ml synthetic RNA is micro-injected into stage 5 or stage 6 oocytes according to methods established in the literature (Gurdon et al (1983) Methods in Enzymol 101.370).

After incubation to allow translation of the mRNAs the oocytes are analysed for expression of the DRG sodium channels via electrophysiological or other methods as described in example 7.

A further method for expression of functional sodium channels involves the nuclear injection of a *Xenopus* oocyte protein expression vector such as pOEV (Pfaff et al., Anal. Biochem. 188, 192-195 (1990)) which allows cloned DNA to be transcribed and translated directly in the oocyte. Since proteins translated in oocytes are post-translationally modified according to conserved eukaryotic signals, these cells offer a convenient system for performing structural and functional analyses of cloned genes. pOEV can be used for direct analysis of proteins encoded by cloned cDNAs without preparing mRNA in vitro, simplifying existing protocols for translating proteins in oocytes with a very high translational yield. Transcription of the vector in oocytes is driven by the promoter for the TFIIA gene, which can generate 1-2 ng (per oocyte within 2 days) of stable mRNA template for translation. The vector also contains SP6 and T7 promoters for in vitro transcription to make mRNA and hybridization probes. DNA clones encoding SNS channel transcripts are injected into oocyte nuclei and protein accumulated in the cell over a 2- to 10-day period. The presence of functional protein is then assessed using twin electrode voltage clamp as described in example 7.

Example 5 - Expression of rat and human DRG sodium channel in mammalian cells

In order to be able to establish a mammalian cell expression system capable of producing the sodium channel in a stable bioactive manner, constructs have to be first generated consisting of the cDNA of the channel in the correct vectors suitable for the cell system in which it is desired to express the protein. There are available a range of vectors containing strong promoters which drive expression in mammalian cells.

i/ Transient expression

In order to determine rapidly the bioactivity of a given cDNA it can be introduced directly into cells and resulting protein activity assayed 48-72 hours later. Although this does not result in a cell line which is stably expressing the protein of interest

it does give a quick answer as to the biological activity of the molecule. Specifically, the cDNA representing the human or rat DRG sodium channel is ligated into appropriate vectors (including but not limited to pRc/RSV, pRc/CMV, pcDNA1 (Invitrogen)) using appropriate restriction enzymes such that the resulting construct contains the cDNA in the correct orientation and such that the heterologous promoter can drive expression of the transcription unit. The resulting expression constructs are introduced into appropriate cell lines including but not limited to COS-7 cells (an African Green Monkey Kidney cell line), HEK 293 cells (a human embryonic kidney cell line) and NIH3T3 cells (a murine fibroblastic cell line). The DNA is introduced via standard methods (Sambrook et al 1989 Molecular Cloning Second Edition, Cold Spring Harbour Laboratory Press) including but not limited to calcium phosphate transfection, electroporation or lipofectamine (Gibco) transfection. After the required incubation time at 37°C in a humidified incubator the cells are tested for the presence of an active rat DRG sodium channel using methods described in example 7.

15

ii/Stable expression

The production of a stable expression system has several advantages over transient expression. A clonal cell line can be generated that has a stable phenotype and in which the expression levels of the foreign protein can be characterised and, with some expression systems, controlled. Also, a range of vectors are available which incorporate genes coding for antibiotic resistance, thus allowing the selection of cells transfected with the constructs introduced. Cell lines of this type can be grown in tissue culture and can be frozen down for long-term storage. There are several systems available for accomplishing this e.g. CHO, CV-1, NIH-3T3.

Specifically COS-7 cells can be transfected by lipofection using Lipofectamine (GibcoBRL) in the following manner. For each sample 2×10^6 cells are seeded in a 90mm tissue culture plate the day prior to transfection. These are incubated overnight at 37°C in a CO₂ incubator to give 50-80% confluency the following day. The day of the transfection the following solutions are prepared in sterile 12 x 75mm tubes:

30 Solution A: For each transfection, dilute 10-50µg of DNA into 990µl of serum-free media (Opti-MEM I Reduced Serum Medium GibcoBRL). Solution B: For each transfection.

dilute 50µl of Lipofectamine Reagent into 950µl serum-free medium. The two solutions are combined, mixed gently and incubated at room temp for 45 minutes. During this time the cells are rinsed once with serum-free medium. For each transfection 9ml of serum-free medium is added to the DNA-lipofectamine tubes. This solution is mixed gently and
5 overlaid on the rinsed cells. The plates are incubated for 5 hours at 37°C in a CO₂ incubator. After the incubation the medium is replaced with fresh complete media and the cells returned to the incubator. Cells are assayed for activity 72 hours post transfection as detailed in examples 4 and 7. To ascertain the efficiency of transfection, β-galactosidase in pcDNA3 is transfected alongside the DRG sodium channel cDNA. This control plate is
10 stained for β-galactosidase activity using a chromogenic substrate and the proportion of cells staining calculated. For transient transfection of DRG the cDNA must first be cloned into a eucaryotic expression vector such as pcDNA3 (Invitrogen).

Example 6 - Expression of rat DRG sodium channel in insect cells

15

The baculovirus expression system uses baculovirus such as Autographa californica nuclear polyhedrosis virus (AcNPV) to produce large amounts of target protein in insect cells such as the Sf9 or 21 clonal cell lines derived from *Spodoptera frugiperda* cells. Expression of the highly abundant polyhedrin gene is non-essential in tissue culture
20 and its strong promoter (polh) can be used for the synthesis of foreign gene products (Smith et al 1983 Mol Cell Biol 3,2156-2165). The polyhedrin promoter is maximally expressed very late in infection (20 hours post infection).

A transfer vector, where the rat DRG sodium channel cDNA is cloned downstream of the polh promoter, or another late promoter such as p10, is transfected into
25 insect cells in conjunction with modified AcNPV viral DNA such as but not limited to BaculoGold DNA (PharMingen). The modified DNA contains a lethal mutation and is incapable of producing infectious viral particles after transfection. Co-transfection with a complementing transfer vector such as (but not limited to) pAcYM1 (Matsuura et al 1987 J Gen Virol 68,1233-1250) or pVL1392/3 (InVitrogen) allows the production of viable
30 recombinant virus. Although more than 99% of the resultant virus particles should be derived from plasmid-rescued virus it is desirable to further purify the virus particles by plaque assay. To ensure that the recombinant stock is clonal, a single plaque is picked from

the plaque assay and amplified to produce a recombinant viral stock. Once the recombinant phenotype is verified the viral stock can be used to infect insect cells and express functional rat DRG sodium channel. There are a number of variations in the methodology of baculovirus expression which may give increased expression (O'Reilly et al 1992
5 Baculovirus Expression Vectors: A Laboratory Manual. Oxford University Press). The expression of the rat or human DRG sodium channel is achieved by cloning of the cDNA into pVL1392 and introducing this into Sf21 insect cells.

10 **Example 7 - Electrophysiological characterisation of cloned human and rat DRG sodium channel expression**

Xenopus laevis oocytes are used to express the channel after injection of the mRNA or cDNA in an expression vector. Expression would be transient and thus functional studies would be made at appropriate times after the injections. Comparison
15 with mock-injected oocytes would demonstrate lack of the novel channel as an endogenously expressed characteristic. Standard two electrode voltage clamp (TEVC) techniques as described, for example, in Fraser, Moon & Djamgoz (1993) Electrophysiology of *Xenopus* oocytes: an expression system in molecular neurobiology. In: Electrophysiology: A practical approach. Wallis, D.I., ed. Oxford University Press.
20 Chapter 4 pp. 65-86, would be used to examine the characteristics of responses of ionic currents to changes in the applied membrane potential. Appropriately modified saline media would be used to manipulate the type of ionic currents detectable. The kinetics of activation and inactivation of the sodium current, its ionic selectivity, the effects of changes in ionic concentration of the extracellular medium on its reversal potential, and the
25 sensitivity (or resistance) to TTX would be defining characteristics.

Similar electrophysiological studies would be undertaken to assess the success of functional expression in a permanently or transiently expressing mammalian cell line, but patch clamp methods would be more suitable than TEVC. Whole cell, cell-attached patch, inside-out patch or outside-out patch configurations as described for
30 example by Hamill et al. (1981) Pflugers Arch. 391:85-100 and Fenwick et al. (1982) J. Physiol. 331 599-635 might be used to assess the channel characteristics.

For example, isolated transfected cells (see above) will be voltage-clamped using the whole-cell variant of the patch clamp technique for recording the expressed sodium channel current.

Recordings will be obtained at room temperature (22-24°C). Both external
5 and internal recording solutions will be used to isolate Na⁺ currents as previously described (Lalick et al., Am. J. Physiol. 264:C803-C809, 1992; West et al., Neuron 8:59-70, 1992). External solution (mM): sodium chloride, 65; choline chloride, 50; TEA-Cl, 20, KCl, 1.5; calcium chloride, 1; magnesium chloride, 5; glucose 5; HEPES, 5; at a pH 7.4 and and osmolality of 320. Internal solution (mM):CsF, 90; CsCl, 60; sodium chloride, 10;
10 MgCl, 2;EGTA, 10; HEPES, 10 at pH 7.2 and an osmolarity of 315.

The kinetics and voltage parameters of the expressed sodium channel current will be examined and compared with data existing in the literature. These include current-voltage relationships and peak current amplitude. Cells will be voltage-clamped at -70 mV and depolarizing pulses to 50 mV (at 10 mV increments) will be used to
15 generate currents.

The pharmacology of the expressed sodium channel current will be examined with the Na channel blocker, tetrodotoxin (TTX). To date sodium channels have been classified as TTX-sensitive and TTX-resistant: block by low (1-30 nM) and high (> 1 µM) concentrations of TTX, respectively (Elliot & Elliot, J. Physiol. (Lond.) 463:39-56,
20 1993; Yang et al., J. Neurosci. 12:268-277, 1992; W1992).

The channel is unaffected by concentrations lower than 1 micromolar tetrodotoxin, and is only partially blocked by concentrations as high as 10 micromolar tetrodotoxin.

25 **Example 8 - Production of purified channel**

Using a commercial coupled transcription-translation system, 35-S methionine-labelled protein products of the SNS clone can be generated (see Figure 3). The size of the resulting protein when assessed by SDS-polyacrylamide gel electrophoresis
30 confirms the predicted size of the protein deduced by DNA sequencing. The system used

is the Promega TNT system (Promega Technical Bulletin 126 1993). The experiment is carried out precisely according to the protocol provided (see Figure 3).

Example 9 - Use of rat or human sodium channel in screening assays

5

Cell lines expressing the cloned sodium channels could be used to determine the effects of drugs on the ability of the channels to pass sodium ions across the cell membranes, e.g to block the channels or to enhance their opening. Since the channel activation is voltage dependent, depolarising conditions will be required for observation of baseline activity that would be modified by drug actions. Depolarisation could be achieved by for example raising extracellular potassium ion concentration to 20 or 40 mM, or by repeated electrical pulses. Detection of the activation of sodium conducting channels could be achieved by flux of radiolabelled sodium ions, guanine or by reporter gene activation leading to for example a colour change or to fluorescence of a light emitting protein.

15 Subsequent confirmation of the effectiveness of the drug action on sodium channel activity would require electrophysiological studies similar to those described above.

Example 10 - In vitro influx assays

- 20 1. $^{22}\text{Na}^+$ influx assay: A modified assay has been adapted from methods reported by Tamkun and Catterall, Mol Pharm. 19:78, (1981). Oocytes or cells expressing the sodium channel gene are suspended in a buffer containing 0.13 M sodium chloride, 5 mM KCl, 0.8 mM MgSO_4 , 50 mM HEPES-Tris (pH 7.4), and 5.5 mM glucose. Aliquots of the
- 25 cell suspension are added a buffer containing $^{22}\text{NaCl}$ (1.3 $\mu\text{Ci/ml}$, New England Nuclear, Boston, MA), 0.128 M choline chloride, 2.66 mM sodium chloride, 5.4 mM KCl, 0.8 mM MgSO_4 , 50 mM HEPES-Tris (pH 7.4), 5 mM ouabain, 1mg/ml bovine serum albumin, and 5.5 mM glucose and then incubated at 37 oC for 20 sec in either the presence or absence of 100 μM veratridine (Sigma Chemical Co., St Louis, MO). The influx assay is stopped by
- 30 the addition of 3 ml of ice-cold wash buffer containing 0.163 M sodium chloride, 0.8 mM MgSO_4 , 1.8 mM CaCl_2 , 50 mM HEPES-Tris (pH 7.4) and 1mg/ml bovine serum albumin,

collected on a glass fiber filter (Whatman GF/C), and washed twice with 3 ml of wash buffer. Radioactive incorporation is determined by with a gammacounter. The specific tetrodotoxin-resistant influx is measured by the difference in $^{22}\text{Na}^+$ uptake in the absence or the presence of $10\text{ }\mu\text{M}$ transmethrin or $1\text{ }\mu\text{M}$ (+) trans allethrin. The

5 tetrodotoxin-sensitive influx is measured by the difference in $^{22}\text{Na}^+$ uptake in the absence or the presence of $1\text{ }\mu\text{M}$ tetrodotoxin (Sigma Chemical Co., St Louis, MO).

Guanidine influx: Another assay is modified from the method described by Reith, Eur. J. Pharmacol. 188:33 (1990). In this assay sodium ions are substituted with guanidinium ions. Oocytes or cells are washed twice with a buffer containing 4.74 mM

10 KCl , 1.25 mM CaCl_2 , 1.2 mM KH_2PO_4 , 1.18 mM MgSO_4 , 22 mM HEPES (pH 7.2), 22 mM choline chloride and 11 mM glucose. The oocytes or cells are suspended in the same buffer containing $250\text{ }\mu\text{M}$ guanidine for 5 min at $19\text{--}25\text{ }^\circ\text{C}$. An aliquot of ^{14}C -labelled guanidine hydrochloride ($30\text{--}50\text{ mCi/mmol}$ supplied by New England Nuclear, Boston, MA) is added in the absence or presence of $10\text{ }\mu\text{M}$ veratridine, and the mixture is

15 incubated for 3 min. The uptake reaction is stopped by filtration through Whatman GF/F filters and followed by 2 5 ml washes with ice-cold 0.9% saline. Radioactive incorporation is determined by scintillation counting.

Example 11

20 In order to measure the expression of sodium channels in in vitro systems, as well as to analyse distribution and relative level of expression in vivo, and to attempt to block function, polyclonal and monoclonal antibodies will be generated to peptide and protein fragments derived from SNS protein sequence shown in Figure 1.

25 a) Immunogens

Glutathione-sulphotransferase (GST) - fusion proteins will be constructed (Smith and Johnson Gene 67:31-40 (1988)) using PGEX vectors obtained from Pharmacia. Fusion proteins including both intracellular and extracellular loops with little homology

30 with known sodium channels other than SNS-B will be produced. One such method involves subcloning of fragments into pGex-5X3 or pGEX 4t-2 to produce in-frame fusion

proteins encoding extracellular, intracellular or C-terminal domains as shown in detailed maps in Figure 4. The pGEX fusion vectors are transformed into *E. coli* XL-1 blue cells or other appropriate cells grown in the presence of ampicillin. After the cultures have reached an optical density of $OD_{600} > 0.5$, fusion protein synthesis is induced by the addition of 100 micromolar IPTG, and the cultures further incubated for 1- 4 hours. The cells are harvested by centrifugation and washed in ice cold phosphate buffered saline. The resuting pellet (dissolved in 300 microlitres PBS from each 50 ml culture) is then sonicated on ice using a 2mm diameter probe, and the lysed cells microfuged to remove debris. 50 microlitres of glutathione-agarose beads are then added to each pellet, and after gentle mixing for 2 minutes at room temperature, the beads are washed by successive spins in PBS. The washed beads are then boiled in Laemmli gel sample buffer, and applied to 10% polyacrylamide SDS gels. Material migrating at the predicted molecular weight is identified on the gel by brief staining with coomassie blue, and comparison with molecular weight markers. This material is then electroeluted from the gel and used as an immunogen as described below.

b) Antibody production

Female Balb/c mice are immunised intraperitoneally with 1-100 micrograms of GST fusion protein emulsified in Freund's complete adjuvant. After 4 weeks, the animals will be further immunised with fusion proteins (1-100 micrograms) emulsified in Freund's incomplete adjuvant. Four weeks later, the animals will be immunised intraperitoneally with a further 1-100 micrograms of GST fusion protein emulsified with Freund's incomplete adjuvant. Seven days later, the animals will be bled, and their serum assessed for the production of antibodies to the immunogen by the following screen; (protocols for the production of rabbit polyclonal serum are the same, except that all injections are subcutaneous, and 10 times as much immunogen is used. Polyclonal rabbit serum are isolated from ear-vein bleeds.)

Serial ten-fold dilutions of the sera (1:100 to 1: 1000,000) in phosphate buffered saline (PBS) containing 0.5% NP-40 and 1% normal goat serum will be applied to 4% paraformaldehyde-fixed 10 micron sections of neonatal rat spinal cord previously treated with 10% goat serum in PBS. After overnight incubation, the sections are washed in

PBS, and further incubated in the dark with 1;200 FITC-conjugated F(ab)2 fragment of goat anti-mouse antibodies for 2 hours in PBS containing 1% normal goat serum. The sections are further washed in PBS, mounted in Citifluor, and examined by fluorescence microscopy. Those sera that show specific staining of laminar II in the spinal cord will be retained, and the mice generating such antibodies subsequently used for the production of monoclonal antibodies. Three weeks later, mice producing useful antibodies are immunised with GST-fusion proteins without adjuvant. After 3 days, the animals are killed, their spleens removed, and the lymphocytes fused with the thymidine kinase-negative myeloma line NS0 or equivalent, using polyethylene glycol. The fused cells from each experiment are grown up in 3 x 24 well plates in the presence of DMEM medium containing 10% fetal calf serum and hypoxanthine, aminopterin and thymidine (HAT) medium to kill the myeloma cells (Kohler and Milstein, Eur. J. Immunol 6, 511-519 (1976)). The tissue culture supernatants from wells containing hybridomas are further screened by immunofluorescence as described above, and cells from positive wells cloned by limiting dilution. Antibody from the positive testing cloned hybridomas is then used to Western blot extracts of rat dorsal root ganglia, to determine if the antibody recognises a band of size approximately 200,000, confirming the specificity of the monoclonal antibody for the SNS sodium channel. Those antibodies directed against extracellular domains that test positive by both of these criteria will then be assessed for function blocking activity in electrophysiological tests of sodium channel function (see example 7), and in screens relying on ion flux or dye-based assays in cells lines expressing sodium channel (see examples 9 and 10).

Example 12 - Cell-type distribution of expression

25

In situ hybridization demonstrates the presence of SNS in a subset of sensory neurons. An SNS fragment between positions 1740 and 1960 was sub-cloned into pGem4z, and DIG-UTP labeled sense or antisense cRNA generated. Sample preparation, hybridization, and visualization of in situ hybridization with alkaline phosphatase conjugated anti-DIG antibodies was carried out exactly as described in Schaeren-Wimers N. and Gerfin-Moser A. Histochemistry 100, 431-440 (1993).

30

Example 13 - Electrophysiological Properties of the Rat DRG Sodium Channel Expressed in *Xenopus oocytes*

pBluescript SK plasmid containing DNA encoding the SNS sodium channel was digested to position -21 upstream of the initiator methionine using a commercially available kit (Erase a base system, Promega, Madison, Wisconsin, USA). The linearized and digested plasmid was cut with Kpn1 and subcloned into an oocyte expression vector pSp64GL (Sma-Kpn1) sites. pSP64GL is derived from pSP64.T pSP64.T was cut with Sma1-EcoR1, blunt-ended with Klenow enzyme, and recircularized. Part of the pGem 72 (+) polylinker (Sma1-Kpn1-EcoR1-Xho1) was ligated into the blunt-ended Bgl II site of pSP64.T. This vector with an altered polylinker for DNA inserts (Sma1-Kpn1-EcoR1-Xho1) and linearization (Sal1-Xba 1-BamH1) was named pSP64GL. The resulting plasmid was linearized with Xba1, and cRNA transcribed with SP6 polymerase using 1 mM 7-methylGppG.

cRNA (70 ng) was injected into *Xenopus* oocytes 7-14 days before recording; immature, stage IV oocytes were chosen cause of their smaller diameter and therefore capacitance. Oocytes were impaled with 3M KCl electrodes ($\leq 1\text{M}\Omega$) and perfused at 3-4 ml per minute with modified Ringer solution containing 115 mM NaCl, 2.5 mM KCl, 10 mM HEPES, 1.8 mM MgCl_2 , and 1 mM CaCl_2 , pH 7.2, at temperature of 19.5 - 20.5 °C. Digital leak subtraction of two electrode voltage-clamp current records was carried out using as leak currents produced by hyperpolarizing pulses of the same amplitude as the test depolarizing commands. Oocytes in which leak commands elicited time-dependent currents were discarded. Averages of 10 records were used for both test and leak.

Inward currents were evoked by depolarizing, in 10 mV steps, from -60 mV to a command potential of -20 to +40 mV in 10 mV steps and from -80 mV to a command potential of -30 to +2- mV in oocytes injected with sodium channel cRNA. Current traces are blanked for the first 1.5 ms from the onset of the voltage step to delete the capacity transients for clarity. The peak current is reached at the same command voltage for the two holding potentials, but is slightly smaller from -60 mV because of steady-state inactivation.

The effects of 50% or 100% replacement of external Na^+ by N-methyl-D-glucosamine on the sodium channel current were elicited by stepping the depolarizing currents given to the oocyte from -60 to +1 mV. Data were fitted with the equation $h_x =$

$1/(1 + \exp((V - V_{50})/k))$, where V is the prepulse potential, V_{50} the potential of 50% inactivation and the k the slope factor (best squares fit). The effect of TTX (10 μ M and 100 μ M) on the peak Na^+ current (test pulse from -60 to +20 mV) was also determined. The effect was quickly reversible upon washout.

5 After a minimum incubation of 7 days from cRNA injection, step depolarizations to potentials positive to -30mV elicited inward currents which peaked between +10 and +20 mV with an average maximum amplitude of 164 ± 72 nA (from -60 mV holding potential, $n = 13$) and a reversal potential of $+35.5 \pm 2.2$ mV ($n = 10$). The inward current was reversed by total replacement of Na^+ in the external medium with an impermeant cation
10 (N-methyl-D-glucosamine). The current's reversal potential was shifted in 50% Na^+ by 13.7 ± 3.2 mV in the hyperpolarizing direction ($n = 3$; predicted value for a Na^+ -selective channel, 17.5 mV). The inactivation produced by a 1s prepulse was half-maximal at -30.0 ± 1.3 mV (slope factor 14.0 ± 1.7 mV, $n = 5$).

TTX had no effect at nanomolar concentrations, and produced only a $19.1 \pm 8.3\%$
15 reduction at 10 μ M, $n = 3$). The estimated half-maximal inhibitory concentration (IC_{50}) was 59.6 ± 10.1 μ M TTX.

The local anesthetic lignocaine was also weakly inhibitory, producing a maximum block of $41.7 \pm 5.4\%$ at 1 mM on the peak current elicited by depolarizing pulses from -60 mV to +10 mV (1 every min; $n = 3$), whereas under the same conditions 100 μ M phenytoin
20 had no effect.

A similarity with the TTX-insensitive Na^+ current of DRG neurons was the effectiveness and rank order of Pb^{2+} versus Cd^{2+} in reducing peak Na^+ currents ($-63.9 \pm 18.1\%$ for Pb^{2+} versus $-24.4 \pm 7.9\%$ for Cd^{2+} at 50 μ M and 100 μ M, respectively; $n = 3$, $P = 0.0189$). The electrophysiological and pharmacological characteristics of the oocyte
25 expressed DRG sodium channel are thus similar to the properties of the sensory neuron TTX-insensitive channel, given the constraints of expression in an oocyte system. In oocytes expressing the DRG sodium channel, the peak of the I/V plot occurred at a more depolarized potential than that of the DRG TTX-insensitive current, despite a similar reversal potential. This difference may reflect the absence of the accessory $\beta 1$ subunit
30 found in DRG, which is known to shift activation to more negative potentials when

expressed with the subunit of other Na⁺ channels. In addition, splice variants that exhibit an activation threshold more negative to SNS sodium channel may shift activation to the more negative potentials observed in sensory neurons.

5 **Example 14 - Distribution of DRG Sodium Channel in Neonatal and Adult Rat Tissues and Cell Lines**

Northern blot and reverse transcriptase-polymerase chain reaction (RT-PCR) were used to examine neonatal and adult rat tissues for expression of the DRG sodium channel
10 messenger RNA.

Random primed ³²P-labeled DNA Pst -Acc1 fragment probes (50 ng, specific activity 2 x 10⁹ c.p.m. per µg DNA) from interdomain region 1 (nucleotide position 1,478-1,892) of the SNS sodium channel nucleic acid sequence were used to probe total RNA extracted from tissues. The following tissues and cell lines were tested: central nervous
15 system and non-neuronal tissues from neonatal rats; peripheral nervous tissue including neonatal Schwann cells and sympathetic neurons, as well as C6 glioma, human embryonal carcinoma line N-tera-2 and N-tera-2 neuro, rat sensory neuron-derived lines ND7 and ND8, and human neuroblastomas SMS-KCN and PC12 cells grown in the presence of NGF; adult rat tissue including pituitary, superior cervical ganglia, coeliac ganglia,
20 trigeminal mesencephalic nucleus, vas deferens, bladder, ileum and DRG of adult animals treated with capsaicin (50 mg/kg) at birth and neonatal DRG control. Total RNA (10 µg) or 25 µg of RNA from tissues apart from superior cervical ganglion sample (10 µg) and capsaicin-treated adult rat DRG (5µg) were northern blotted.

Total RNA was separated on 1.2% agarose-formaldehyde gels, and capillary blotted
25 onto Hibond-N filters (Amersham). The amounts of RNA on the blot were roughly equivalent, as judged by ethidium bromide staining of ribosomal RNA and by hybridization with the ubiquitously expressed L-27 ribosomal protein transcripts. Filters were prehybridized in 50% formamide, 5 x SSC containing 0.5% sodium dodecyl sulfate, 5 x Denhardt's solution, 100 µg/ml boiled sonicated salmon sperm DNA (average size 300
30 bp), 10 µg/ml poly-U and 10 µg/ml poly-C at 45°C for 6h. After 36 hours hybridization in the same conditions using 10⁷ c.p.m. per ml hybridization probe, the filters were briefly washed in 2 x SSC at room temperature, then twice with 2 x SSC with 0.5% SDS at 68°C

for 15 min, followed by a 20 min wash in 0.5% SDS, 0.2 x SSC at 68°C. The filters were autoradiographed overnight or for 4 days on autoradiography film (Kodak X-omat).

For RT-PCR experiments, 10 µg total RNA from neonatal rat tissues (spleen, liver, kidney, lung, intestine, muscle, heart, superior cervical ganglia, spinal cord, brain stem, hippocampus, cerebellum, cortex and dorsal root ganglia), or 2 µg total RNA from control or capsaicin-treated rat DRG or DRG neurons in culture were treated with DNase I and extracted with acidic phenol to remove genomic DNA.

cDNA was synthesized with Superscript reverse transcriptase using oligo dT(12-18) primers and purified on Qiagen 5 tips. Polymerase chain reaction (PCR) was used to amplify cDNA (35 cycles, 94°C, 1 min; 55°C, 1 min; and 72°C, 1 min), and products separated on agarose gels before staining with ethidium bromide. L-27 primers (Ninkina et al. (1983) Nucleic Acids Res. 21, 3175-3182) were added to the PCR reaction 5 cycles after the start of the reaction with the DRG sodium channel specific primers which comprised

5'-CAGCTTCGCTCAGAAGTATCT-3' (SEQ ID NO: 9) and
5'-TTCTCGCCGTTCCACACGGAGA-3' (SEQ ID NO: 10).

Transcription of mRNA coding for the DRG sodium channel could not be detected in any non-neuronal tissues or in the central nervous system using northern blots or reverse transcription of mRNA and the polymerase chain reaction. Sympathetic neurons from the superior cervical ganglion and Schwann cell-containing sciatic nerve preparations, as well as several neuronal cell lines were also negative. However, total RNA extracts from neonatal and adult rat DRG gave a strong signal of size about 7kb on northern blots. These data suggest that the DRG sodium channel is not expressed only in early development.

RT-PCR of oligo dT-primed cDNA from various tissues using DRG sodium channel primers and L-27 ribosomal protein primer showed the presence of DRG sodium channel transcripts in DRG tissue only.

RT-PCR was also performed on DRG-sodium channel and L-27 transcripts from DRG neurons cultured and treated with capsaicin (overnight 10 µM) or dissected from neonatal animals treated with capsaicin (50 mg/kg on 2 consecutive days, followed by DRG isolation 5 days later. The signal from the L-27 probe was the same in capsaicin-treated cell cultures or animals as compared with controls that were not treated with

capsaicin. There was a significant diminution in the DRG sodium channel signal from capsaicin-treated cultures or animals as compared with controls. Control PCR reactions without reverse transcriptase treatment were also done to control for contaminating genomic DNA.

- 5 When neonatal rats were treated with capsaicin and total adult DRG RNA subsequently examined by northern blotting, the signal was substantially reduced, suggesting that the DRG sodium channel transcript is expressed selectively by capsaicin-sensitive (predominantly nociceptive) neurons. These data were confirmed by RT-PCR experiments on both cultures of DRG neurons, and in whole animal studies.

10

Example 15 - Distribution of DRG sodium channel in rat tissue by *in situ* hybridization

- In situ* hybridization was used to examine the expression of the DRG sodium
15 channel transcripts at the single-cell level in both adult trigeminal ganglia and neonatal and adult rat DRG.

- A SNS sodium channel PCR fragment of interdomain region I between positions 1,736 and 1,797 of the SNS sodium channel nucleic acid sequence was subcloned into pGem3Z (Promega, Madison, Wisconsin, USA) and digoxigenin (DIG)-UTP (Boehringer-Mannheim, Germany) labeled sense or antisense cRNA generated using SP6 or T7
20 polymerase, respectively. Sample preparation, hybridization and visualization of *in situ* hybridization with alkaline phosphatase conjugated anti-DIG antibodies was carried out as described in Schaeren-Wimers, et al., A. (1993) Histochemistry 100: 431-440, with the following modifications. Frozen tissue sections (10 μ M-thick) of neonatal rat lumbar
25 DRG, and adult trigeminal ganglion neurons were fixed for 10 min in phosphate buffered saline (PBS) containing 4% paraformaldehyde. Sections were acetylated in 0.1M triethanolamine, 0.25% acetic anhydride for 10 min. Prehybridization was carried out in 50% formamide, 4 x SSC, 100 μ g/ml boiled and sonicated ssDNA, 50 μ g/ml yeast tRNA, 2 x Denhardt's solution at room temperature for 1 h. Hybridization was carried out
30 overnight in the same buffer at 65°C. Probe concentration was 50 ng/ml. Sections were washed in 2 x SSC for 30 min at 72°C for 1 hr and twice in 0.1 SSC for 30 min at 72°C

before visualization at room temperature with anti-digoxigenin alkaline phosphatase conjugated antibodies. The same sections were then stained with mouse monoclonal antibody RT97 which is specific for neurofilaments found in large diameter neurons.

Subsets of sensory neurons from both tissues showed intense signals with a DRG sodium channel-specific probe. Combined immunohistochemistry with the large-diameter neuron-specific monoclonal antibody RT97 and the DRG sodium channel specific probe showed that most of the large diameter neurons did not express the DRG sodium channel transcript. Small diameter neurons were stained with the DRG sodium channel specific probe but not the large diameter neurons.

10

Example 16 - Site Directed Mutagenesis of SNS Sodium Channel - TTX Sensitivity

The SNS sodium channel is 65% homologous to the tetrodotoxin-insensitive cardiac sodium channel. A number of residues that line the channel atrium have been implicated in tetrodotoxin binding. The amino acid sequence of the SNS sodium channel exhibits sequence identity to other tetrodotoxin-sensitive sodium channels in 7 out of 9 such residues. One difference is a conservative substitution at D(905)E. A single residue (C-357) has been shown to play a critical role in tetrodotoxin binding to the sodium channel. In the SNS sodium channel, a hydrophilic serine is found at this position, whereas other sodium channels that are sensitive to TTX have phenylalanine in this position.

Site-directed mutagenesis using standard techniques and primers having the sequence TGACGCAGGACTCCTGGGAGCGCC (SEQ ID NO: 31) was used to substitute phenylalanine for serine at position 357 in the SNS sodium channel. The mutated SNS sodium channel, when expressed in *Xenopus* oocytes produces voltage-gated currents similar in amplitude and time course to the native channel. However, sensitivity to TTX is restored to give an IC_{50} of 2.5 nM (± 0.4 , $n = 5$), similar to other voltage-gated sodium channels that have aromatic residues at the equivalent position. The table below shows IC_{50} for SNS sodium channel, and the rat brain iia, muscle type 1, and cardiac tetrodotoxin-insensitive sodium channels.

30

TTX Sensitivity

Sodium Channel	ss1 domain	ss2 domain	IC_{50}
----------------	------------	------------	-----------

Rat brain iia	FRLM	TQDFWENLY	18 nM
muscle type 1	FRLM	TQDYWENLY	40 nM
cardiac TTXi	FRLM	TQDCWERLY	950 nM
SNS	FRLM	TQDSWERLY	60 micromolar
SNS mutant	FRLM	TQDFWERLY	2.5 nM

FRLM - SEQ ID NO: 11; TQDFWENLY - SEQ ID NO: 12;

TQDYWENLY - SEQ ID NO: 13; TQDCWERLY - SEQ ID NO:14;

TQDSWERLY - SEQ ID NO: 15; TQDFWERLY - SEQ ID NO:16

5

Example 18

Polyclonal antibodies were raised in rabbits against the following peptides derived from the SNS sodium channel protein amino acid sequence:

- Peptide 1 TQDSWER (SEQ ID NO:17)
- 10 Peptide 2 GSTDDNRSPQSDPYN (SEQ ID NO: 18)
- Peptide 3 SPKENHGDFI (SEQ ID NO: 19)
- Peptide 4 PNHNGSRGN (SEQ ID NO: 20)

The peptides were conjugated to Keyhole limpet heocyanin (KLH) and injected repeatedly into rabbits. Sera from the rabbits was treated by Western blotting. Several sera showed

15 positive results indicating the presence of antibodies specific for the peptide in the sera.

References

- Catterall W .A. (1992) *Physiol. Rev.* 72, S4-S47.
- 20 Cohen S.A. and Barchi R.L. (1993) *Int. Rev. Cytology* 137c, 55-103.
- Hodgkin A.L. and Huxley A.F. (1952) *J. Physiol.* 116, 473-496.
- Hille B. (1991) *Ionic channels in excitable membranes* (Sinauer Sunderland MA)
- Jeftjina S. (1994) *Brain Res.* 639, 125-134.
- Kohler G. and Milstein C. (1976) *Eur J. Immunol* 6, 511-519
- 25 Lewin B. (1995) *Genes V* Oxford University Press, Oxford.
- Melton D. et al. (1984) *Nucleic Acids Res.* 12, 7035

- Nowycky M. (1993) in Sensory Neurons (Ed Scott S.) OUP, Oxford.
- Omri G. and Meir H. (1990) J. Membrane Biol. 115, 13-29
- Pearce R.J. and Duchen M.R. (1994) Neuroscience 63, 1041-1056
- Pfaff SL; Tamkun-MM; Taylor-WL (1990) Anal-Biochem. 1990 188 192-195
- 5 Schaeren-Wimers N. and Gerfin-Moser A. (1993) Histochemistry 100, 431-440.
- Smith D.B. and Johnson K.S. (1988) Gene 67, 31-40.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

5 (i) APPLICANT:
 (A) NAME: University College London
 (B) STREET: Gower Street
 (C) CITY: London
 10 (E) COUNTRY: England
 (F) POSTAL CODE (ZIP): WC1E 6BT
 (ii) TITLE OF INVENTION: Ion Channel
 15 (iii) NUMBER OF SEQUENCES: 31
 (iv) COMPUTER READABLE FORM:
 (A) MEDIUM TYPE: Floppy disk
 (B) COMPUTER: IBM PC compatible
 20 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(2) INFORMATION FOR SEQ ID NO:1:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 6524 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 30 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

35 (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 204..6077

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

50	TAGCTTGCTT CTGCTAATGC TACCCCAGGC CTTTAGACAG AGAACAGATG GCAGATGGAG	60
45	TTTCTTATTG CCATGCGCAA ACGCTGAGCC CACCTCATGA TCCCGGACCC CATGGTTTTC	120
	AGTAGACAAC CTGGGCTAAG AAGAGATCTC CGACCTTATA GAGCAGCAAA GAGTGTAAT	180
	TCTTCCCCAA GAAGAATGAG AAG ATG GAG CTC CCC TTT GCG TCC GTG GGA	230
	Met Glu Leu Pro Phe Ala Ser Val Gly	
50	1 5	
	ACT ACC AAT TTC AGA CGG TTC ACT CCA GAG TCA CTG GCA GAG ATC GAG	278
	Thr Thr Asn Phe Arg Arg Phe Thr Pro Glu Ser Leu Ala Glu Ile Glu	
	10 15 20 25	
55	AAG CAG ATT GCT GCT CAC CGC GCA GCC AAG AAG GCC AGA ACC AAG CAC	326
	Lys Gln Ile Ala Ala His Arg Ala Ala Lys Lys Ala Arg Thr Lys His	
	30 35 40	
60	AGA GGA CAG GAG GAC AAG GGC GAG AAG CCC AGG CCT CAG CTG GAC TTG	374
	Arg Gly Gln Glu Asp Lys Gly Glu Lys Pro Arg Pro Gln Leu Asp Leu	
	45 50 55	

	AAA GAC TGT AAC CAG CTG CCC AAG TTC TAT GGT GAG CTC CCA GCA GAA	422
	Lys Asp Cys Asn Gln Leu Pro Lys Phe Tyr Gly Glu Leu Pro Ala Glu	
	60 65 70	
5	CTG GTC GGG GAG CCC CTG GAG GAC CTA GAC CCT TTC TAC AGC ACA CAC	470
	Leu Val Gly Glu Pro Leu Glu Asp Leu Asp Pro Phe Tyr Ser Thr His	
	75 80 85	
10	CGG ACA TTC ATG GTG TTG AAT AAA AGC AGG ACC ATT TCC AGA TTC AGT	518
	Arg Thr Phe Met Val Leu Asn Lys Ser Arg Thr Ile Ser Arg Phe Ser	
	90 95 100 105	
15	GCC ACT TGG GCC CTG TGG CTC TTC AGT CCC TTC AAC CTG ATC AGA AGA	566
	Ala Thr Trp Ala Leu Trp Leu Phe Ser Pro Phe Asn Leu Ile Arg Arg	
	110 115 120	
20	ACA GCC ATC AAA GTG TCT GTC CAT TCC TGG TTC TCC ATA TTC ATC ACC	614
	Thr Ala Ile Lys Val Ser Val His Ser Trp Phe Ser Ile Phe Ile Thr	
	125 130 135	
25	ATC ACT ATT TTG GTC AAC TGC GTG TGC ATG ACC CGA ACT GAT CTT CCA	662
	Ile Thr Ile Leu Val Asn Cys Val Cys Met Thr Arg Thr Asp Leu Pro	
	140 145 150	
30	GAG AAA GTC GAG TAC GTC TTC ACT GTC ATT TAC ACC TTC GAG GCT CTG	710
	Glu Lys Val Glu Tyr Val Phe Thr Val Ile Tyr Thr Phe Glu Ala Leu	
	155 160 165	
35	ATT AAG ATA CTG GCA AGA GGG TTT TGT CTA AAT GAG TTC ACT TAT CTT	758
	Ile Lys Ile Leu Ala Arg Gly Phe Cys Leu Asn Glu Phe Thr Tyr Leu	
	170 175 180 185	
40	CGA GAT CCG TGG AAC TGG CTG GAC TTC AGT GTC ATT ACC TTG GCG TAT	806
	Arg Asp Pro Trp Asn Trp Leu Asp Phe Ser Val Ile Thr Leu Ala Tyr	
	190 195 200	
45	GTG GGT GCA GCG ATA GAC CTC CGA GGA ATC TCA GGC CTG CGG ACA TTC	854
	Val Gly Ala Ala Ile Asp Leu Arg Gly Ile Ser Gly Leu Arg Thr Phe	
	205 210 215	
50	CGA GTT CTC AGA GCC CTG AAA ACT GTT TCT GTG ATC CCA GGA CTG AAG	902
	Arg Val Leu Arg Ala Leu Lys Thr Val Ser Val Ile Pro Gly Leu Lys	
	220 225 230	
55	GTC ATC GTG GGA GCC CTG ATC CAC TCA GTG AGG AAG CTG GCC GAC GTG	950
	Val Ile Val Gly Ala Leu Ile His Ser Val Arg Lys Leu Ala Asp Val	
	235 240 245	
60	ACT ATC CTC ACA GTC TTC TGC CTG AGC GTC TTC GCC TTG GTG GGC CTG	998
	Thr Ile Leu Thr Val Phe Cys Leu Ser Val Phe Ala Leu Val Gly Leu	
	250 255 260 265	
65	CAG CTC TTT AAG GGG AAC CTT AAG AAC AAA TGC ATC AGG AAC GGA ACA	1046
	Gln Leu Phe Lys Gly Asn Leu Lys Asn Lys Cys Ile Arg Asn Gly Thr	
	270 275 280	
70	GAT CCC CAC AAG GCT GAC AAC CTC TCA TCT GAA ATG GCA GAA TAC GTC	1094
	Asp Pro His Lys Ala Asp Asn Leu Ser Ser Glu Met Ala Glu Tyr Val	
	285 290 295	
75	TCC ATC AAG CCT GGT ACT ACG GAT CCC TTA CTG TGC GGC AAT GGG TCT	1142
	Ser Ile Lys Pro Gly Thr Thr Asp Pro Leu Leu Cys Gly Asn Gly Ser	
	300 305 310	

	GAT GCT GGT CAC TGC CCT GGA GGC TAT GTC TGC CTG AAA ACT CCT GAC	1190
	Asp Ala Gly His Cys Pro Gly Gly Tyr Val Cys Leu Lys Thr Pro Asp	
	315 320 325	
5	AAC CCG GAT TTT AAC TAC ACC AGC TTT GAT TCC TTT GCG TGG GCA TTC	1238
	Asn Pro Asp Phe Asn Tyr Thr Ser Phe Asp Ser Phe Ala Trp Ala Phe	
	330 335 340 345	
10	CTC TCA CTG TTC CGC CTC ATG ACG CAG GAC TCC TGG GAG CGC CTG TAC	1286
	Leu Ser Leu Phe Arg Leu Met Thr Gln Asp Ser Trp Glu Arg Leu Tyr	
	350 355 360	
15	CAG CAG ACA CTC CGG GCT TCT GGG AAA ATG TAC ATG GTC TTT TTC GTG	1334
	Gln Gln Thr Leu Arg Ala Ser Gly Lys Met Tyr Met Val Phe Phe Val	
	365 370 375	
20	CTG GTT ATT TTC CTT GGA TCG TTC TAC CTG GTC AAT TTG ATC TTG GCC	1382
	Leu Val Ile Phe Leu Gly Ser Phe Tyr Leu Val Asn Leu Ile Leu Ala	
	380 385 390	
25	GTG GTC ACC ATG GCG TAT GAA GAG CAG AGC CAG GCA ACA ATT GCA GAA	1430
	Val Val Thr Met Ala Tyr Glu Glu Gln Ser Gln Ala Thr Ile Ala Glu	
	395 400 405	
30	ATC GAA GCC AAG GAA AAA AAG TTC CAG GAA GCC CTT GAG GTG CTG CAG	1478
	Ile Glu Ala Lys Glu Lys Phe Gln Glu Ala Leu Glu Val Leu Gln	
	410 415 420 425	
35	AAG GAA CAG GAG GTG CTG GCA GCC CTG GGG ATT GAC ACG ACC TCG CTC	1526
	Lys Glu Gln Glu Val Leu Ala Ala Leu Gly Ile Asp Thr Thr Ser Leu	
	430 435 440	
40	CAG TCC CAC AGT GGA TCA CCC TTA GCC TCC AAA AAC GCC AAT GAG AGA	1574
	Gln Ser His Ser Gly Ser Pro Leu Ala Ser Lys Asn Ala Asn Glu Arg	
	445 450 455	
45	AGA CCC AGG GTG AAA TCA AGG GTG TCA GAG GGC TCC ACG GAT GAC AAC	1622
	Arg Pro Arg Val Lys Ser Arg Val Ser Glu Gly Ser Thr Asp Asp Asn	
	460 465 470	
50	AGG TCA CCC CAA TCT GAC CCT TAC AAC CAG CGC AGG ATG TCT TTC CTA	1670
	Arg Ser Pro Gln Ser Asp Pro Tyr Asn Gln Arg Arg Met Ser Phe Leu	
	475 480 485	
55	GGC CTG TCT TCA GGA AGA CGC AGG GCT AGC CAC GGC AGT GTG TTC CAC	1718
	Gly Leu Ser Ser Gly Arg Arg Arg Ala Ser His Gly Ser Val Phe His	
	490 495 500 505	
60	TTC CGA GCG CCC AGC CAA GAC ATC TCA TTT CCT GAC GGG ATC ACC CCT	1766
	Phe Arg Ala Pro Ser Gln Asp Ile Ser Phe Pro Asp Gly Ile Thr Pro	
	510 515 520	
65	GAT GAT GGG GTC TTT CAC GGA GAC CAG GAA AGC CGT CGA GGT TCC ATA	1814
	Asp Asp Gly Val Phe His Gly Asp Gln Glu Ser Arg Arg Gly Ser Ile	
	525 530 535	
70	TTG CTG GGC AGG GGT GCT GGG CAG ACA GGT CCA CTC CCC AGG AGC CCA	1862
	Leu Leu Gly Arg Gly Ala Gly Gln Thr Gly Pro Leu Pro Arg Ser Pro	
	540 545 550	
75	CTG CCT CAG TCC CCC AAC CCT GGC CGT AGC CAT GGA GAA GAG GGA CAG	1910
	Leu Pro Gln Ser Pro Asn Pro Gly Arg Arg His Gly Glu Glu Gly Gln	
	555 560 565	

-53-

	CTC	GGA	GTG	CCC	ACT	GGT	GAG	CTT	ACC	GCT	GGA	GCG	CCT	GAA	GGC	CCG	1958
	Leu	Gly	Val	Pro	Thr	Gly	Glu	Leu	Thr	Ala	Gly	Ala	Pro	Glu	Gly	Pro	
	570					575					580					585	
5	GCA	CTG	CAC	ACT	ACA	GGG	CAG	AAG	AGC	TTC	CTG	TCT	GCG	GGC	TAC	TTG	2006
	Ala	Leu	His	Thr	Thr	Gly	Gln	Lys	Ser	Phe	Leu	Ser	Ala	Gly	Tyr	Leu	
					590					595					600		
10	AAC	GAA	CCT	TTC	CGA	GCA	CAG	AGG	GCG	ATG	AGC	GTT	GTC	AGT	ATC	ATG	2054
	Asn	Glu	Pro	Phe	Arg	Ala	Gln	Arg	Ala	Met	Ser	Val	Val	Ser	Ile	Met	
				605					610					615			
15	ACT	TCT	GTC	ATT	GAG	GAG	CTT	GAA	GAG	TCT	AAG	CTG	AAG	TGC	CCA	CCC	2102
	Thr	Ser	Ile	Ile	Glu	Glu	Leu	Glu	Glu	Ser	Lys	Leu	Lys	Cys	Pro	Pro	
			620					625					630				
20	TGC	TTG	ATC	AGC	TTC	GCT	CAG	AAG	TAT	CTG	ATC	TGG	GAG	TGC	TGC	CCC	2150
	Cys	Leu	Ile	Ser	Phe	Ala	Gln	Lys	Tyr	Leu	Ile	Trp	Glu	Cys	Cys	Pro	
		635				640						645					
25	AAG	TGG	AGG	AAG	TTC	AAG	ATG	GCG	CTG	TTC	GAG	CTG	GTG	ACT	GAC	CCC	2198
	Lys	Trp	Arg	Lys	Phe	Lys	Met	Ala	Leu	Phe	Glu	Leu	Val	Thr	Asp	Pro	
		650				655					660					665	
30	TTC	GCA	GAG	CTT	ACC	ATC	ACC	CTC	TGC	ATC	GTG	GTG	AAC	ACC	GTC	TTC	2246
	Phe	Ala	Glu	Leu	Thr	Ile	Thr	Leu	Cys	Ile	Val	Val	Asn	Thr	Val	Phe	
					670					675					680		
35	CAA	GCC	GGC	AAC	ATT	GTC	TTC	ACC	GTG	TTT	TTC	ACA	ATG	GAG	ATG	GCC	2342
	Gln	Ala	Gly	Asn	Ile	Val	Phe	Thr	Val	Phe	Phe	Thr	Met	Glu	Met	Ala	
			700					705					710				
40	TTC	AAG	ATC	ATT	GCC	TTC	GAC	CCC	TAC	TAT	TAC	TTC	CAG	AAG	AAG	TGG	2390
	Phe	Lys	Ile	Ile	Ala	Phe	Asp	Pro	Tyr	Tyr	Tyr	Phe	Gln	Lys	Lys	Trp	
		715				720						725					
45	AAT	ATC	TTC	GAC	TGT	GTC	ATC	GTC	ACC	GTG	AGC	CTT	CTG	GAG	CTG	AGT	2438
	Asn	Ile	Phe	Asp	Cys	Val	Ile	Val	Thr	Val	Ser	Leu	Leu	Glu	Leu	Ser	
		730				735					740					745	
50	GCA	TCC	AAG	AAG	GGC	AGC	CTG	TCT	GTG	CTC	CGT	ACC	TTA	CGC	TTG	CTG	2486
	Ala	Ser	Lys	Lys	Gly	Ser	Leu	Ser	Val	Leu	Arg	Thr	Leu	Arg	Leu	Leu	
					750					755					760		
55	CGG	GTC	TTC	AAG	CTG	GCC	AAG	TCC	TGG	CCC	ACC	CTG	AAC	ACC	CTC	ATC	2534
	Arg	Val	Phe	Lys	Leu	Ala	Lys	Ser	Trp	Pro	Thr	Leu	Asn	Thr	Leu	Ile	
				765					770					775			
60	AAG	ATC	ATC	GGG	AAC	TCA	GTG	GGG	GCC	CTG	GGC	AAC	CTG	ACC	TTT	ATC	2582
	Lys	Ile	Ile	Gly	Asn	Ser	Val	Gly	Ala	Leu	Gly	Asn	Leu	Thr	Phe	Ile	
			780					785					790				
65	CTG	GCC	ATC	ATC	GTC	TTC	ATC	TTC	GCC	CTG	GTC	GGA	AAG	CAG	CTT	CTC	2630
	Leu	Ala	Ile	Ile	Val	Phe	Ile	Phe	Ala	Leu	Val	Gly	Lys	Gln	Leu	Leu	
		795				800						805					
70	TCA	GAG	GAC	TAC	GGG	TGC	CGC	AAG	GAC	GGC	GTC	TCC	GTG	TGG	AAC	GGC	2678
	Ser	Glu	Asp	Tyr	Gly	Cys	Arg	Lys	Asp	Gly	Val	Ser	Val	Trp	Asn	Gly	
		810				815					820					825	

	GAG AAG CTC CGC TGG CAC ATG TGT GAC TTC TTC CAT TCC TTC CTG GTC	2726
	Glu Lys Leu Arg Trp His Met Cys Asp Phe Phe His Ser Phe Leu Val	
	830 835 840	
5	GTC TTC CGA ATC CTC TGC GGG GAG TGG ATC GAG AAC ATG TGG GTC TGC	2774
	Val Phe Arg Ile Leu Cys Gly Glu Trp Ile Glu Asn Met Trp Val Cys	
	845 850 855	
10	ATG GAG GTC AGC CAG AAA TCC ATC TGC CTC ATC CTC TTC TTG ACT GTG	2822
	Met Glu Val Ser Gln Lys Ser Ile Cys Leu Ile Leu Phe Leu Thr Val	
	860 865 870	
15	ATG GTG CTG GGC AAC CTA GTG GTG CTC AAC CTT TTC ATC GCT TTA CTG	2870
	Met Val Leu Gly Asn Leu Val Val Leu Asn Leu Phe Ile Ala Leu Leu	
	875 880 885	
20	CTG AAC TCC TTC AGC GCG GAC AAC CTC ACG GCT CCA GAG GAT GAC GGG	2918
	Leu Asn Ser Phe Ser Ala Asp Asn Leu Thr Ala Pro Glu Asp Asp Gly	
	890 895 900 905	
25	GAG GTG AAC AAC TTG CAG TTA GCA CTG GCC AGG ATC CAG GTA CTT GGC	2966
	Glu Val Asn Asn Leu Gln Leu Ala Leu Ala Arg Ile Gln Val Leu Gly	
	910 915 920	
30	CAT CGG GCC AGC AGG GCC AGC GCC AGT TAC ATC AGC AGC CAC TGC CGA	3014
	His Arg Ala Ser Arg Ala Ser Ala Ser Tyr Ile Ser Ser His Cys Arg	
	925 930 935	
35	TTC CAC TGG CCC AAG GTG GAG ACC CAG CTG GGC ATG AAG CCC CCA CTC	3062
	Phe His Trp Pro Lys Val Glu Thr Gln Leu Gly Met Lys Pro Pro Leu	
	940 945 950	
40	ACC AGC TCA GAG GCC AAG AAC CAC ATT GCC ACT GAT GCT GTC AGT GCT	3110
	Thr Ser Ser Glu Ala Lys Asn His Ile Ala Thr Asp Ala Val Ser Ala	
	955 960 965	
45	GCA GTG GGG AAC CTG ACA AAG CCA GCT CTC AGT AGC CCC AAG GAG AAC	3158
	Ala Val Gly Asn Leu Thr Lys Pro Ala Leu Ser Ser Pro Lys Glu Asn	
	970 975 980 985	
50	CAC GGG GAC TTC ATC ACT GAT CCC AAC GTG TGG GTC TCT GTG CCC ATT	3206
	His Gly Asp Phe Ile Thr Asp Pro Asn Val Trp Val Ser Val Pro Ile	
	990 995 1000	
55	GCT GAG GGG GAA TCT GAC CTC GAC GAG CTC GAG GAA GAT ATG GAG CAG	3254
	Ala Glu Gly Glu Ser Asp Leu Asp Glu Leu Glu Glu Asp Met Glu Gln	
	1005 1010 1015	
60	GCT TCG CAG AGC TCC TGG CAG GAA GAG GAC CCC AAG GGA CAG CAG GAG	3302
	Ala Ser Gln Ser Ser Trp Gln Glu Glu Asp Pro Lys Gly Gln Gln Glu	
	1020 1025 1030	
65	CAG TTG CCA CAA GTC CAA AAG TGT GAA AAC CAC CAG GCA GCC AGA AGC	3350
	Gln Leu Pro Gln Val Gln Lys Cys Glu Asn His Gln Ala Ala Arg Ser	
	1035 1040 1045	
70	CCA GCC TCC ATG ATG TCC TCT GAG GAC CTG GCT CCA TAC CTG GGT GAG	3398
	Pro Ala Ser Met Met Ser Ser Glu Asp Leu Ala Pro Tyr Leu Gly Glu	
	1050 1055 1060 1065	
75	AGC TGG AAG AGG AAG GAT AGC CCT CAG GTC CCT GCC GAG GGA GTG GAT	3446
	Ser Trp Lys Arg Lys Asp Ser Pro Gln Val Pro Ala Glu Gly Val Asp	
	1070 1075 1080	

	GAC	ACG	AGC	TCC	TCT	GAG	GGC	AGC	ACG	GTG	GAC	TGC	CCG	GAC	CCA	GAG	3494
	Asp	Thr	Ser	Ser	Ser	Glu	Gly	Ser	Thr	Val	Asp	Cys	Pro	Asp	Pro	Glu	
				1085					1090					1095			
5	GAA	ATC	CTG	AGG	AAG	ATC	CCC	GAG	CTG	GCA	CAT	GAC	CTG	GAC	GAG	CCC	3542
	Glu	Ile	Leu	Arg	Lys	Ile	Pro	Glu	Leu	Ala	His	Asp	Leu	Asp	Glu	Pro	
			1100					1105					1110				
10	GAT	GAC	TGT	TTC	AGA	GAA	GGC	TGC	ACT	CGC	CGC	TGT	CCC	TGC	TGC	AAC	3590
	Asp	Asp	Cys	Phe	Arg	Glu	Gly	Cys	Thr	Arg	Arg	Cys	Pro	Cys	Cys	Asn	
			1115				1120					1125					
15	GTG	AAT	ACT	AGC	AAG	TCT	CCT	TGG	GCC	ACA	GGC	TGG	CAG	GTG	CGC	AAG	3638
	Val	Asn	Thr	Ser	Lys	Ser	Pro	Trp	Ala	Thr	Gly	Trp	Gln	Val	Arg	Lys	
	1130					1135					1140					1145	
20	ACC	TGC	TAC	CGC	ATC	GTG	GAG	CAC	AGC	TGG	TTT	GAG	AGT	TTC	ATC	ATC	3686
	Thr	Cys	Tyr	Arg	Ile	Val	Glu	His	Ser	Trp	Phe	Glu	Ser	Phe	Ile	Ile	
				1150						1155					1160		
25	TTC	ATG	ATC	CTG	CTC	AGC	AGT	GGA	GCG	CTG	GCC	TTT	GAG	GAT	AAC	TAC	3734
	Phe	Met	Ile	Leu	Leu	Ser	Ser	Gly	Ala	Leu	Ala	Phe	Glu	Asp	Asn	Tyr	
			1165					1170					1175				
30	CTG	GAA	GAG	AAA	CCC	CGA	GTG	AAG	TCC	GTG	CTG	GAG	TAC	ACT	GAC	CGA	3782
	Leu	Glu	Glu	Lys	Pro	Arg	Val	Lys	Ser	Val	Leu	Glu	Tyr	Thr	Asp	Arg	
			1180					1185					1190				
35	GTG	TTC	ACC	TTC	ATC	TTC	GTC	TTT	GAG	ATG	CTG	CTC	AAG	TGG	GTA	GCC	3830
	Val	Phe	Thr	Phe	Ile	Phe	Val	Phe	Glu	Met	Leu	Leu	Lys	Trp	Val	Ala	
		1195					1200					1205					
40	TAT	GGC	TTC	AAA	AAG	TAT	TTC	ACC	AAT	GCC	TGG	TGC	TGG	CTG	GAC	TTC	3878
	Tyr	Gly	Phe	Lys	Lys	Tyr	Phe	Thr	Asn	Ala	Trp	Cys	Trp	Leu	Asp	Phe	
	1210					1215				1220						1225	
45	CTC	ATT	GTG	AAC	ATC	TCC	CTG	ACA	AGC	CTC	ATA	GCG	AAG	ATC	CTT	GAG	3926
	Leu	Ile	Val	Asn	Ile	Ser	Leu	Thr	Ser	Leu	Ile	Ala	Lys	Ile	Leu	Glu	
				1230						1235					1240		
50	TAT	TCC	GAC	GTG	GCG	TCC	ATC	AAA	GCC	CTT	CGG	ACT	CTC	CGT	GCC	CTC	3974
	Tyr	Ser	Asp	Val	Ala	Ser	Ile	Lys	Ala	Leu	Arg	Thr	Leu	Arg	Ala	Leu	
				1245				1250					1255				
55	CGA	CCG	CTG	CGG	GCT	CTG	TCT	CGA	TTC	GAA	GGC	ATG	AGG	GTA	GTG	GTG	4022
	Arg	Pro	Leu	Arg	Ala	Leu	Ser	Arg	Phe	Glu	Gly	Met	Arg	Val	Val	Val	
			1260					1265					1270				
60	GAT	GCC	CTC	GTG	GGC	GCC	ATC	CCC	TCC	ATC	ATG	AAC	GTC	CTC	CTC	GTC	4070
	Asp	Ala	Leu	Val	Gly	Ala	Ile	Pro	Ser	Ile	Met	Asn	Val	Leu	Leu	Val	
		1275					1280					1285					
65	TGC	CTC	ATC	TTC	TGG	CTC	ATC	TTC	AGC	ATC	ATG	GGC	GTG	AAC	CTC	TTC	4118
	Cys	Leu	Ile	Phe	Trp	Leu	Ile	Phe	Ser	Ile	Met	Gly	Val	Asn	Leu	Phe	
	1290					1295					1300					1305	
70	GCC	GGG	AAA	TTT	TCG	AAG	TGC	GTC	GAC	ACC	AGA	AAT	AAC	CCA	TTT	TCC	4166
	Ala	Gly	Lys	Phe	Ser	Lys	Cys	Val	Asp	Thr	Arg	Asn	Asn	Pro	Phe	Ser	
				1310						1315					1320		
75	AAC	GTG	AAT	TCG	ACG	ATG	GTG	AAT	AAC	AAG	TCC	GAG	TGT	CAC	AAT	CAA	4214
	Asn	Val	Asn	Ser	Thr	Met	Val	Asn	Asn	Lys	Ser	Glu	Cys	His	Asn	Gln	
				1325						1330					1335		

-56-

	AAC AGC ACC GGC CAC TTC TTC TGG GTC AAC GTC AAA GTC AAC TTC GAC	4262
	Asn Ser Thr Gly His Phe Phe Trp Val Asn Val Lys Val Asn Phe Asp	
	1340 1345 1350	
5	AAC GTC GCT ATG GGC TAC CTC GCA CTT CTT CAG GTG GCA ACC TTC AAA	4310
	Asn Val Ala Met Gly Tyr Leu Ala Leu Leu Gln Val Ala Thr Phe Lys	
	1355 1360 1365	
10	GGC TGG ATG GAC ATA ATG TAT GCA GCT GTT GAT TCC GGA GAG ATC AAC	4358
	Gly Trp Met Asp Ile Met Tyr Ala Ala Val Asp Ser Gly Glu Ile Asn	
	1370 1375 1380 1385	
15	AGT CAG CCT AAC TGG GAG AAC AAC TTG TAC ATG TAC CTG TAC TTC GTC	4406
	Ser Gln Pro Asn Trp Glu Asn Asn Leu Tyr Met Tyr Leu Tyr Phe Val	
	1390 1395 1400	
20	GTT TTC ATC ATT TTC GGT GGC TTC TTC ACG CTG AAT CTC TTT GTT GGG	4454
	Val Phe Ile Ile Phe Gly Gly Phe Phe Thr Leu Asn Leu Phe Val Gly	
	1405 1410 1415	
25	GTC ATA ATC GAC AAC TTC AAC CAA CAG AAA AAA AAG CTA GGA GGC CAG	4502
	Val Ile Ile Asp Asn Phe Asn Gln Gln Lys Lys Lys Leu Gly Gly Gln	
	1420 1425 1430	
30	GAC ATC TTC ATG ACA GAA GAG CAG AAG AAG TAC TAC AAT GCC ATG AAG	4550
	Asp Ile Phe Met Thr Glu Glu Gln Lys Lys Tyr Tyr Asn Ala Met Lys	
	1435 1440 1445	
35	AAG CTG GGC TCC AAG AAA CCC CAG AAG CCC ATC CCA CGG CCC CTG AAT	4598
	Lys Leu Gly Ser Lys Lys Pro Gln Lys Pro Ile Pro Arg Pro Leu Asn	
	1450 1455 1460 1465	
40	AAG TAC CAA GGC TTC GTG TTT GAC ATC GTG ACC AGG CAA GCC TTT GAC	4646
	Lys Tyr Gln Gly Phe Val Phe Asp Ile Val Thr Arg Gln Ala Phe Asp	
	1470 1475 1480	
45	ATC ATC ATC ATG GTT CTC ATC TGC CTC AAC ATG ATC ACC ATG ATG GTG	4694
	Ile Ile Ile Met Val Leu Ile Cys Leu Asn Met Ile Thr Met Met Val	
	1485 1490 1495	
50	GAG ACC GAC GAG CAG GGC GAG GAG AAG ACG AAG GTT CTG GGC AGA ATC	4742
	Glu Thr Asp Glu Gln Gly Glu Glu Lys Thr Lys Val Leu Gly Arg Ile	
	1500 1505 1510	
55	AAC CAG TTC TTT GTG GCC GTC TTC ACG GGC GAG TGT GTG ATG AAG ATG	4790
	Asn Gln Phe Phe Val Ala Val Phe Thr Gly Glu Cys Val Met Lys Met	
	1515 1520 1525	
60	TTC GCC CTG CGA CAG TAC TAC TTC ACC AAC GGC TGG AAC GTG TTC GAC	4838
	Phe Ala Leu Arg Gln Tyr Tyr Phe Thr Asn Gly Trp Asn Val Phe Asp	
	1530 1535 1540 1545	
65	TTC ATA GTG GTG ATC CTG TCC ATT GGG AGT CTG CTG TTT TCT GCA ATC	4886
	Phe Ile Val Val Ile Leu Ser Ile Gly Ser Leu Leu Phe Ser Ala Ile	
	1550 1555 1560	
70	CTT AAG TCA CTG GAA AAC TAC TTC TCC CCG ACG CTC TTC CGG GTC ATC	4934
	Leu Lys Ser Leu Glu Asn Tyr Phe Ser Pro Thr Leu Phe Arg Val Ile	
	1565 1570 1575	
75	CGT CTG GCC AGG ATC GGC CGC ATC CTC AGG CTG ATC CGA GCA GCC AAG	4982
	Arg Leu Ala Arg Ile Gly Arg Ile Leu Arg Leu Ile Arg Ala Ala Lys	
	1580 1585 1590	

-57-

	GGG	ATT	CGC	ACG	CTG	CTC	TTC	GCC	CTC	ATG	ATG	TCC	CTG	CCC	GCC	CTC	5030
	Gly	Ile	Arg	Thr	Leu	Leu	Phe	Ala	Leu	Met	Met	Ser	Leu	Pro	Ala	Leu	
	1595						1600					1605					
5	TTC	AAC	ATC	GGC	CTC	CTC	CTC	TTC	CTC	GTC	ATG	TTC	ATC	TAC	TCC	ATC	5078
	Phe	Asn	Ile	Gly	Leu	Leu	Leu	Phe	Leu	Val	Met	Phe	Ile	Tyr	Ser	Ile	
	1610					1615					1620					1625	
10	TTC	GGC	ATG	GCC	AGC	TTC	GCT	AAC	GTC	GTG	GAC	GAG	GCC	GGC	ATC	GAC	5126
	Phe	Gly	Met	Ala	Ser	Phe	Ala	Asn	Val	Val	Asp	Glu	Ala	Gly	Ile	Asp	
					1630					1635					1640		
15	GAC	ATG	TTC	AAC	TTC	AAG	ACC	TTT	GGC	AAC	AGC	ATG	CTG	TGC	CTG	TTC	5174
	Asp	Met	Phe	Asn	Phe	Lys	Thr	Phe	Gly	Asn	Ser	Met	Leu	Cys	Leu	Phe	
				1645					1650					1655			
20	CAG	ATC	ACC	ACC	TCG	GCC	GGC	TGG	GAC	GGC	CTC	CTC	AGC	CCC	ATC	CTC	5222
	Gln	Ile	Thr	Thr	Ser	Ala	Gly	Trp	Asp	Gly	Leu	Leu	Ser	Pro	Ile	Leu	
				1660				1665					1670				
25	AAC	ACG	GGG	CCT	CCC	TAC	TGC	GAC	CCC	AAC	CTG	CCC	AAC	AGC	AAC	GGC	5270
	Asn	Thr	Gly	Pro	Pro	Tyr	Cys	Asp	Pro	Asn	Leu	Pro	Asn	Ser	Asn	Gly	
		1675					1680					1685					
30	TCC	CGG	GGG	AAC	TGC	GGG	AGC	CCG	GCG	GTG	GGC	ATC	ATC	TTC	TTC	ACC	5318
	Ser	Arg	Gly	Asn	Cys	Gly	Ser	Pro	Ala	Val	Gly	Ile	Ile	Phe	Phe	Thr	
	1690					1695					1700					1705	
35	ACC	TAC	ATC	ATC	ATC	TCC	TTC	CTC	ATC	GTG	GTC	AAC	ATG	TAC	ATC	GCA	5366
	Thr	Tyr	Ile	Ile	Ile	Ser	Phe	Leu	Ile	Val	Val	Asn	Met	Tyr	Ile	Ala	
					1710					1715					1720		
40	GTG	ATT	CTG	GAG	AAC	TTC	AAC	GTA	GCC	ACC	GAG	GAG	AGC	ACG	GAG	CCC	5414
	Val	Ile	Leu	Glu	Asn	Phe	Asn	Val	Ala	Thr	Glu	Glu	Ser	Thr	Glu	Pro	
				1725				1730						1735			
45	CTG	AGC	GAG	GAC	GAC	TTC	GAC	ATG	TTC	TAT	GAG	ACC	TGG	GAG	AAG	TTC	5462
	Leu	Ser	Glu	Asp	Asp	Phe	Asp	Met	Phe	Tyr	Glu	Thr	Trp	Glu	Lys	Phe	
			1740					1745					1750				
50	GAC	CCG	GAG	GCC	ACC	CAG	TTC	ATT	GCC	TTT	TCT	GCC	CTC	TCA	GAC	TTC	5510
	Asp	Pro	Glu	Ala	Thr	Gln	Phe	Ile	Ala	Phe	Ser	Ala	Leu	Ser	Asp	Phe	
		1755					1760					1765					
55	GCG	GAC	ACG	CTC	TCC	GGC	CCT	CTT	AGA	ATC	CCC	AAA	CCC	AAC	CAG	AAT	5558
	Ala	Asp	Thr	Leu	Ser	Gly	Pro	Leu	Arg	Ile	Pro	Lys	Pro	Asn	Gln	Asn	
	1770					1775					1780					1785	
60	ATA	TTA	ATC	CAG	ATG	GAC	CTG	CCG	TTG	GTC	CCC	GGG	GAT	AAG	ATC	CAC	5606
	Ile	Leu	Ile	Gln	Met	Asp	Leu	Pro	Leu	Val	Pro	Gly	Asp	Lys	Ile	His	
					1790					1795					1800		
65	TGT	CTG	GAC	ATC	CTT	TTT	GCC	TTC	ACA	AAG	AAC	GTC	TTG	GGA	GAA	TCC	5654
	Cys	Leu	Asp	Ile	Leu	Phe	Ala	Phe	Thr	Lys	Asn	Val	Leu	Gly	Glu	Ser	
				1805				1810						1815			
70	GGG	GAG	TTG	GAC	TCC	CTG	AAG	ACC	AAT	ATG	GAA	GAG	AAG	TTT	ATG	GCG	5702
	Gly	Glu	Leu	Asp	Ser	Leu	Lys	Thr	Asn	Met	Glu	Glu	Lys	Phe	Met	Ala	
			1820					1825					1830				
75	ACC	AAT	CTC	TCC	AAA	GCA	TCC	TAT	GAA	CCA	ATA	GCC	ACC	ACC	CTC	CGG	5750
	Thr	Asn	Leu	Ser	Lys	Ala	Ser	Tyr	Glu	Pro	Ile	Ala	Thr	Thr	Leu	Arg	
		1835					1840					1845					

-58-

TGG AAG CAG GAA GAC CTC TCA GCC ACA GTC ATT CAA AAG GCC TAC CGG 5798
 Trp Lys Gln Glu Asp Leu Ser Ala Thr Val Ile Gln Lys Ala Tyr Arg
 1850 1855 1860 1865
 5 AGC TAC ATG CTG CAC CGC TCC TTG ACA CTC TCC AAC ACC CTG CAT GTG 5846
 Ser Tyr Met Leu His Arg Ser Leu Thr Leu Ser Asn Thr Leu His Val
 1870 1875 1880
 10 CCC AGG GCT GAG GAG GAT GGC GTG TCA CTT CCC GGG GAA GGC TAC ATT 5894
 Pro Arg Ala Glu Glu Asp Gly Val Ser Leu Pro Gly Glu Gly Tyr Ile
 1885 1890 1895
 15 ACA TTC ATG GCA AAC AGT GGA CTC CCG GAC AAA TCA GAA ACT GCC TCT 5942
 Thr Phe Met Ala Asn Ser Gly Leu Pro Asp Lys Ser Glu Thr Ala Ser
 1900 1905 1910
 GCT ACG TCT TTC CCG CCA TCC TAT GAC AGT GTC ACC AGG GGC CTG AGT 5990
 Ala Thr Ser Phe Pro Pro Ser Tyr Asp Ser Val Thr Arg Gly Leu Ser
 1915 1920 1925
 20 GAC CGG GCC AAC ATT AAC CCA TCT AGC TCA ATG CAA AAT GAA GAT GAG 6038
 Asp Arg Ala Asn Ile Asn Pro Ser Ser Ser Met Gln Asn Glu Asp Glu
 1930 1935 1940 1945
 25 GTC GCT GCT AAG GAA GGA AAC AGC CCT GGA CCT CAG TGAAGGCACT 6084
 Val Ala Ala Lys Glu Gly Asn Ser Pro Gly Pro Gln
 1950 1955
 30 CAGGCATGCA CAGGGCAGGT TCCAATGTCT TTCTCTGCTG TACTAACTCC TTCCCTCTGG 6144
 AGGTGGCACC AACCTCCAGC CTCCACCAAT GCATGTCACT GGTCATGGTG TCAGAACTGA 6204
 ATGGGGACAT CCTTGAGAAA GCCCCACCC CAATAGGAAT CAAAAGCCAA GGATACTCCT 6264
 35 CCATTCTGAC GTCCCTTCCG AGTTCCAGAG AGATGTCATT GCTCCCTTCT GTTTGTGACC 6324
 AGAGACGTGA TTCACCAACT TCTCGGAGCC AGAGACACAT AGCAAAGACT TTTCTGCTGG 6384
 TGTCGGGCAG TCTTAGAGAA GTCACGTAGG GGTGGTACT GAGAATTAGG GTTTCATGA 6444
 40 CTGCATGCTC ACAGCTGCCG GACAATACCT GTGAGTCGGC CATTAAAATT AATATTTTAA 6504
 AAGTTAAAAA AAAAAAAAAA 6524

45

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1957 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

50

(ii) MOLECULE TYPE: protein

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Glu Leu Pro Phe Ala Ser Val Gly Thr Thr Asn Phe Arg Arg Phe
 1 5 10 15
 60 Thr Pro Glu Ser Leu Ala Glu Ile Glu Lys Gln Ile Ala Ala His Arg
 20 25 30
 Ala Ala Lys Lys Ala Arg Thr Lys His Arg Gly Gln Glu Asp Lys Gly
 35 40 45

-59-

	Glu	Lys	Pro	Arg	Pro	Gln	Leu	Asp	Leu	Lys	Asp	Cys	Asn	Gln	Leu	Pro	
	50						55					60					
5	Lys	Phe	Tyr	Gly	Glu	Leu	Pro	Ala	Glu	Leu	Val	Gly	Glu	Pro	Leu	Glu	
	65					70					75					80	
	Asp	Leu	Asp	Pro	Phe	Tyr	Ser	Thr	His	Arg	Thr	Phe	Met	Val	Leu	Asn	
					85					90					95		
10	Lys	Ser	Arg	Thr	Ile	Ser	Arg	Phe	Ser	Ala	Thr	Trp	Ala	Leu	Trp	Leu	
				100					105					110			
	Phe	Ser	Pro	Phe	Asn	Leu	Ile	Arg	Arg	Thr	Ala	Ile	Lys	Val	Ser	Val	
15			115					120					125				
	His	Ser	Trp	Phe	Ser	Ile	Phe	Ile	Thr	Ile	Thr	Ile	Leu	Val	Asn	Cys	
		130					135					140					
20	Val	Cys	Met	Thr	Arg	Thr	Asp	Leu	Pro	Glu	Lys	Val	Glu	Tyr	Val	Phe	
	145					150					155					160	
	Thr	Val	Ile	Tyr	Thr	Phe	Glu	Ala	Leu	Ile	Lys	Ile	Leu	Ala	Arg	Gly	
					165					170					175		
25	Phe	Cys	Leu	Asn	Glu	Phe	Thr	Tyr	Leu	Arg	Asp	Pro	Trp	Asn	Trp	Leu	
				180					185					190			
	Asp	Phe	Ser	Val	Ile	Thr	Leu	Ala	Tyr	Val	Gly	Ala	Ala	Ile	Asp	Leu	
30			195					200					205				
	Arg	Gly	Ile	Ser	Gly	Leu	Arg	Thr	Phe	Arg	Val	Leu	Arg	Ala	Leu	Lys	
		210					215					220					
35	Thr	Val	Ser	Val	Ile	Pro	Gly	Leu	Lys	Val	Ile	Val	Gly	Ala	Leu	Ile	
	225					230					235					240	
	His	Ser	Val	Arg	Lys	Leu	Ala	Asp	Val	Thr	Ile	Leu	Thr	Val	Phe	Cys	
					245					250					255		
40	Leu	Ser	Val	Phe	Ala	Leu	Val	Gly	Leu	Gln	Leu	Phe	Lys	Gly	Asn	Leu	
				260					265					270			
	Lys	Asn	Lys	Cys	Ile	Arg	Asn	Gly	Thr	Asp	Pro	His	Lys	Ala	Asp	Asn	
45			275					280					285				
	Leu	Ser	Ser	Glu	Met	Ala	Glu	Tyr	Val	Ser	Ile	Lys	Pro	Gly	Thr	Thr	
		290					295					300					
50	Asp	Pro	Leu	Leu	Cys	Gly	Asn	Gly	Ser	Asp	Ala	Gly	His	Cys	Pro	Gly	
	305					310					315					320	
	Gly	Tyr	Val	Cys	Leu	Lys	Thr	Pro	Asp	Asn	Pro	Asp	Phe	Asn	Tyr	Thr	
					325					330					335		
55	Ser	Phe	Asp	Ser	Phe	Ala	Trp	Ala	Phe	Leu	Ser	Leu	Phe	Arg	Leu	Met	
				340					345					350			
	Thr	Gln	Asp	Ser	Trp	Glu	Arg	Leu	Tyr	Gln	Gln	Thr	Leu	Arg	Ala	Ser	
60			355					360					365				
	Gly	Lys	Met	Tyr	Met	Val	Phe	Phe	Val	Leu	Val	Ile	Phe	Leu	Gly	Ser	
		370					375					380					

-60-

	Phe	Tyr	Leu	Val	Asn	Leu	Ile	Leu	Ala	Val	Val	Thr	Met	Ala	Tyr	Glu	385	390	395	400
5	Glu	Gln	Ser	Gln	Ala	Thr	Ile	Ala	Glu	Ile	Glu	Ala	Lys	Glu	Lys	Lys	405	410	415	
	Phe	Gln	Glu	Ala	Leu	Glu	Val	Leu	Gln	Lys	Glu	Gln	Glu	Val	Leu	Ala	420	425	430	
10	Ala	Leu	Gly	Ile	Asp	Thr	Thr	Ser	Leu	Gln	Ser	His	Ser	Gly	Ser	Pro	435	440	445	
	Leu	Ala	Ser	Lys	Asn	Ala	Asn	Glu	Arg	Arg	Pro	Arg	Val	Lys	Ser	Arg	450	455	460	
15	Val	Ser	Glu	Gly	Ser	Thr	Asp	Asp	Asn	Arg	Ser	Pro	Gln	Ser	Asp	Pro	465	470	475	480
20	Tyr	Asn	Gln	Arg	Arg	Met	Ser	Phe	Leu	Gly	Leu	Ser	Ser	Gly	Arg	Arg	485	490	495	
	Arg	Ala	Ser	His	Gly	Ser	Val	Phe	His	Phe	Arg	Ala	Pro	Ser	Gln	Asp	500	505	510	
25	Ile	Ser	Phe	Pro	Asp	Gly	Ile	Thr	Pro	Asp	Asp	Gly	Val	Phe	His	Gly	515	520	525	
	Asp	Gln	Glu	Ser	Arg	Arg	Gly	Ser	Ile	Leu	Leu	Gly	Arg	Gly	Ala	Gly	530	535	540	
30	Gln	Thr	Gly	Pro	Leu	Pro	Arg	Ser	Pro	Leu	Pro	Gln	Ser	Pro	Asn	Pro	545	550	555	560
35	Gly	Arg	Arg	His	Gly	Glu	Glu	Gly	Gln	Leu	Gly	Val	Pro	Thr	Gly	Glu	565	570	575	
	Leu	Thr	Ala	Gly	Ala	Pro	Glu	Gly	Pro	Ala	Leu	His	Thr	Thr	Gly	Gln	580	585	590	
40	Lys	Ser	Phe	Leu	Ser	Ala	Gly	Tyr	Leu	Asn	Glu	Pro	Phe	Arg	Ala	Gln	595	600	605	
	Arg	Ala	Met	Ser	Val	Val	Ser	Ile	Met	Thr	Ser	Val	Ile	Glu	Glu	Leu	610	615	620	
45	Glu	Glu	Ser	Lys	Leu	Lys	Cys	Pro	Pro	Cys	Leu	Ile	Ser	Phe	Ala	Gln	625	630	635	640
50	Lys	Tyr	Leu	Ile	Trp	Glu	Cys	Cys	Pro	Lys	Trp	Arg	Lys	Phe	Lys	Met	645	650	655	
	Ala	Leu	Phe	Glu	Leu	Val	Thr	Asp	Pro	Phe	Ala	Glu	Leu	Thr	Ile	Thr	660	665	670	
55	Leu	Cys	Ile	Val	Val	Asn	Thr	Val	Phe	Met	Ala	Met	Glu	His	Tyr	Pro	675	680	685	
	Met	Thr	Asp	Ala	Phe	Asp	Ala	Met	Leu	Gln	Ala	Gly	Asn	Ile	Val	Phe	690	695	700	
60	Thr	Val	Phe	Phe	Thr	Met	Glu	Met	Ala	Phe	Lys	Ile	Ile	Ala	Phe	Asp	705	710	715	720

-61-

Pro Tyr Tyr Tyr Phe Gln Lys Lys Trp Asn Ile Phe Asp Cys Val Ile
 725 730 735
 5 Val Thr Val Ser Leu Leu Glu Leu Ser Ala Ser Lys Lys Gly Ser Leu
 740 745 750
 Ser Val Leu Arg Thr Leu Arg Leu Leu Arg Val Phe Lys Leu Ala Lys
 755 760 765
 10 Ser Trp Pro Thr Leu Asn Thr Leu Ile Lys Ile Ile Gly Asn Ser Val
 770 775 780
 Gly Ala Leu Gly Asn Leu Thr Phe Ile Leu Ala Ile Ile Val Phe Ile
 785 790 795 800
 15 Phe Ala Leu Val Gly Lys Gln Leu Leu Ser Glu Asp Tyr Gly Cys Arg
 805 810 815
 20 Lys Asp Gly Val Ser Val Trp Asn Gly Glu Lys Leu Arg Trp His Met
 820 825 830
 Cys Asp Phe Phe His Ser Phe Leu Val Val Phe Arg Ile Leu Cys Gly
 835 840 845
 25 Glu Trp Ile Glu Asn Met Trp Val Cys Met Glu Val Ser Gln Lys Ser
 850 855 860
 30 Ile Cys Leu Ile Leu Phe Leu Thr Val Met Val Leu Gly Asn Leu Val
 865 870 875 880
 Val Leu Asn Leu Phe Ile Ala Leu Leu Leu Asn Ser Phe Ser Ala Asp
 885 890 895
 35 Asn Leu Thr Ala Pro Glu Asp Asp Gly Glu Val Asn Asn Leu Gln Leu
 900 905 910
 Ala Leu Ala Arg Ile Gln Val Leu Gly His Arg Ala Ser Arg Ala Ser
 915 920 925
 40 Ala Ser Tyr Ile Ser Ser His Cys Arg Phe His Trp Pro Lys Val Glu
 930 935 940
 45 Thr Gln Leu Gly Met Lys Pro Pro Leu Thr Ser Ser Glu Ala Lys Asn
 945 950 955 960
 His Ile Ala Thr Asp Ala Val Ser Ala Ala Val Gly Asn Leu Thr Lys
 965 970 975
 50 Pro Ala Leu Ser Ser Pro Lys Glu Asn His Gly Asp Phe Ile Thr Asp
 980 985 990
 Pro Asn Val Trp Val Ser Val Pro Ile Ala Glu Gly Glu Ser Asp Leu
 995 1000 1005
 55 Asp Glu Leu Glu Glu Asp Met Glu Gln Ala Ser Gln Ser Ser Trp Gln
 1010 1015 1020
 60 Glu Glu Asp Pro Lys Gly Gln Gln Glu Gln Leu Pro Gln Val Gln Lys
 1025 1030 1035 1040
 Cys Glu Asn His Gln Ala Ala Arg Ser Pro Ala Ser Met Met Ser Ser
 1045 1050 1055

-62-

Glu Asp Leu Ala Pro Tyr Leu Gly Glu Ser Trp Lys Arg Lys Asp Ser
 1060 1065 1070
 5 Pro Gln Val Pro Ala Glu Gly Val Asp Asp Thr Ser Ser Ser Glu Gly
 1075 1080 1085
 Ser Thr Val Asp Cys Pro Asp Pro Glu Glu Ile Leu Arg Lys Ile Pro
 1090 1095 1100
 10 Glu Leu Ala His Asp Leu Asp Glu Pro Asp Asp Cys Phe Arg Glu Gly
 1105 1110 1115 1120
 Cys Thr Arg Arg Cys Pro Cys Cys Asn Val Asn Thr Ser Lys Ser Pro
 1125 1130 1135
 15 Trp Ala Thr Gly Trp Gln Val Arg Lys Thr Cys Tyr Arg Ile Val Glu
 1140 1145 1150
 His Ser Trp Phe Glu Ser Phe Ile Ile Phe Met Ile Leu Leu Ser Ser
 1155 1160 1165
 20 Gly Ala Leu Ala Phe Glu Asp Asn Tyr Leu Glu Glu Lys Pro Arg Val
 1170 1175 1180
 Lys Ser Val Leu Glu Tyr Thr Asp Arg Val Phe Thr Phe Ile Phe Val
 1185 1190 1195 1200
 Phe Glu Met Leu Leu Lys Trp Val Ala Tyr Gly Phe Lys Lys Tyr Phe
 1205 1210 1215
 30 Thr Asn Ala Trp Cys Trp Leu Asp Phe Leu Ile Val Asn Ile Ser Leu
 1220 1225 1230
 Thr Ser Leu Ile Ala Lys Ile Leu Glu Tyr Ser Asp Val Ala Ser Ile
 1235 1240 1245
 35 Lys Ala Leu Arg Thr Leu Arg Ala Leu Arg Pro Leu Arg Ala Leu Ser
 1250 1255 1260
 Arg Phe Glu Gly Met Arg Val Val Val Asp Ala Leu Val Gly Ala Ile
 1265 1270 1275 1280
 Pro Ser Ile Met Asn Val Leu Leu Val Cys Leu Ile Phe Trp Leu Ile
 1285 1290 1295
 45 Phe Ser Ile Met Gly Val Asn Leu Phe Ala Gly Lys Phe Ser Lys Cys
 1300 1305 1310
 Val Asp Thr Arg Asn Asn Pro Phe Ser Asn Val Asn Ser Thr Met Val
 1315 1320 1325
 50 Asn Asn Lys Ser Glu Cys His Asn Gln Asn Ser Thr Gly His Phe Phe
 1330 1335 1340
 Trp Val Asn Val Lys Val Asn Phe Asp Asn Val Ala Met Gly Tyr Leu
 1345 1350 1355 1360
 Ala Leu Leu Gln Val Ala Thr Phe Lys Gly Trp Met Asp Ile Met Tyr
 1365 1370 1375
 60 Ala Ala Val Asp Ser Gly Glu Ile Asn Ser Gln Pro Asn Trp Glu Asn
 1380 1385 1390
 Asn Leu Tyr Met Tyr Leu Tyr Phe Val Val Phe Ile Ile Phe Gly Gly
 1395 1400 1405

Phe Phe Thr Leu Asn Leu Phe Val Gly Val Ile Ile Asp Asn Phe Asn
 1410 1415 1420
 5 Gln Gln Lys Lys Lys Leu Gly Gly Gln Asp Ile Phe Met Thr Glu Glu
 1425 1430 1435 1440
 Gln Lys Lys Tyr Tyr Asn Ala Met Lys Lys Leu Gly Ser Lys Lys Pro
 1445 1450 1455
 10 Gln Lys Pro Ile Pro Arg Pro Leu Asn Lys Tyr Gln Gly Phe Val Phe
 1460 1465 1470
 15 Asp Ile Val Thr Arg Gln Ala Phe Asp Ile Ile Ile Met Val Leu Ile
 1475 1480 1485
 Cys Leu Asn Met Ile Thr Met Met Val Glu Thr Asp Glu Gln Gly Glu
 1490 1495 1500
 20 Glu Lys Thr Lys Val Leu Gly Arg Ile Asn Gln Phe Phe Val Ala Val
 1505 1510 1515 1520
 Phe Thr Gly Glu Cys Val Met Lys Met Phe Ala Leu Arg Gln Tyr Tyr
 1525 1530 1535
 25 Phe Thr Asn Gly Trp Asn Val Phe Asp Phe Ile Val Val Ile Leu Ser
 1540 1545 1550
 30 Ile Gly Ser Leu Leu Phe Ser Ala Ile Leu Lys Ser Leu Glu Asn Tyr
 1555 1560 1565
 Phe Ser Pro Thr Leu Phe Arg Val Ile Arg Leu Ala Arg Ile Gly Arg
 1570 1575 1580
 35 Ile Leu Arg Leu Ile Arg Ala Ala Lys Gly Ile Arg Thr Leu Leu Phe
 1585 1590 1595 1600
 Ala Leu Met Met Ser Leu Pro Ala Leu Phe Asn Ile Gly Leu Leu Leu
 1605 1610 1615
 40 Phe Leu Val Met Phe Ile Tyr Ser Ile Phe Gly Met Ala Ser Phe Ala
 1620 1625 1630
 45 Asn Val Val Asp Glu Ala Gly Ile Asp Asp Met Phe Asn Phe Lys Thr
 1635 1640 1645
 Phe Gly Asn Ser Met Leu Cys Leu Phe Gln Ile Thr Thr Ser Ala Gly
 1650 1655 1660
 50 Trp Asp Gly Leu Leu Ser Pro Ile Leu Asn Thr Gly Pro Pro Tyr Cys
 1665 1670 1675 1680
 Asp Pro Asn Leu Pro Asn Ser Asn Gly Ser Arg Gly Asn Cys Gly Ser
 1685 1690 1695
 55 Pro Ala Val Gly Ile Ile Phe Phe Thr Thr Tyr Ile Ile Ile Ser Phe
 1700 1705 1710
 60 Leu Ile Val Val Asn Met Tyr Ile Ala Val Ile Leu Glu Asn Phe Asn
 1715 1720 1725
 Val Ala Thr Glu Glu Ser Thr Glu Pro Leu Ser Glu Asp Asp Phe Asp
 1730 1735 1740

-64-

Met Phe Tyr Glu Thr Trp Glu Lys Phe Asp Pro Glu Ala Thr Gln Phe
 1745 1750 1755 1760

5 Ile Ala Phe Ser Ala Leu Ser Asp Phe Ala Asp Thr Leu Ser Gly Pro
 1765 1770 1775

Leu Arg Ile Pro Lys Pro Asn Gln Asn Ile Leu Ile Gln Met Asp Leu
 1780 1785 1790

10 Pro Leu Val Pro Gly Asp Lys Ile His Cys Leu Asp Ile Leu Phe Ala
 1795 1800 1805

Phe Thr Lys Asn Val Leu Gly Glu Ser Gly Glu Leu Asp Ser Leu Lys
 15 1810 1815 1820

Thr Asn Met Glu Glu Lys Phe Met Ala Thr Asn Leu Ser Lys Ala Ser
 1825 1830 1835 1840

20 Tyr Glu Pro Ile Ala Thr Thr Leu Arg Trp Lys Gln Glu Asp Leu Ser
 1845 1850 1855

Ala Thr Val Ile Gln Lys Ala Tyr Arg Ser Tyr Met Leu His Arg Ser
 1860 1865 1870

25 Leu Thr Leu Ser Asn Thr Leu His Val Pro Arg Ala Glu Glu Asp Gly
 1875 1880 1885

Val Ser Leu Pro Gly Glu Gly Tyr Ile Thr Phe Met Ala Asn Ser Gly
 30 1890 1895 1900

Leu Pro Asp Lys Ser Glu Thr Ala Ser Ala Thr Ser Phe Pro Pro Ser
 1905 1910 1915 1920

35 Tyr Asp Ser Val Thr Arg Gly Leu Ser Asp Arg Ala Asn Ile Asn Pro
 1925 1930 1935

Ser Ser Ser Met Gln Asn Glu Asp Glu Val Ala Ala Lys Glu Gly Asn
 1940 1945 1950

40 Ser Pro Gly Pro Gln
 1955

45 (2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2573 base pairs
 50 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
 55

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 561..2126
 60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CTGGGAGAGA AAGCGTCTCG CCTAGCGACT CCCAGAGCTT TAAGCCGGGA AGGGACAAGC

60

	GTCAGGACAT CTCAGAATCC CGAACCTTCT AGGGAGGGAG GTTCTTACCT CCATGCTTCC	120
	CGTAGGAACC TAATCCCAAT TATTTAGCTG TATTTATAAT ACAAATATG AATGTTAAAT	180
5	GTACAAAATG CTTTCCAGC ATGCCTGCAT CTCCTCCTAG AGTCCTGTTC CCAAGCCCTC	240
	TCTACTCTCA GTACTGTAGA AAAGAAATAA GCTTTACGTG AGAAACCCAG GCACTGGATC	300
10	TTATCCAGGT GCTCACCTCA GAGTCTTTAG TGGGTGTAGC GCTGTGGTAG AGCATTTGGT	360
	TATAGATACA AACCCAGGGC AGGGAGACTG CAGTGGCCAT TCTCTCCAG GCCAGACGTG	420
	CCCTGATCCT TCCCACAGAG ATGAGAAGGC TGGAACCAGA ACACTCAGGT TTTGGCTTCT	480
15	CTTGGGGGAG GAGAGGTAAT CTTGTTACTT TAATAACATC AGTGTGTCCC TCTCCTCTAC	540
	TAGGAGGCCA GGACATCTTC ATG ACA GAA GAG CAG AAG AAG TAC TAC AAT	590
20	Met Thr Glu Glu Gln Lys Lys Tyr Tyr Asn	
	1 5 10	
	GCC ATG AAG AAG CTG GGC TCC AAG AAA CCC CAG AAG CCC ATC CCA CGG	638
	Ala Met Lys Lys Leu Gly Ser Lys Lys Pro Gln Lys Pro Ile Pro Arg	
	15 20 25	
25	CCC CTG AAT AAG TAC CAA GGC TTC GTG TTT GAC ATC GTG ACC AGG CAA	686
	Pro Leu Asn Lys Tyr Gln Gly Phe Val Phe Asp Ile Val Thr Arg Gln	
	30 35 40	
30	GCC TTT GAC ATC ATC ATC ATG GTT CTC ATC TGC CTC AAC ATG ATC ACC	734
	Ala Phe Asp Ile Ile Ile Met Val Leu Ile Cys Leu Asn Met Ile Thr	
	45 50 55	
	ATG ATG GTG GAG ACC GAC GAG CAG GGC GAG GAG AAG ACG AAG GTT CTG	782
35	Met Met Val Glu Thr Asp Glu Gln Gly Glu Glu Lys Thr Lys Val Leu	
	60 65 70	
	GGC AGA ATC AAC CAG TTC TTT GTG GCC GTC TTC ACG GGC GAG TGT GTG	830
	Gly Arg Ile Asn Gln Phe Phe Val Ala Val Phe Thr Gly Glu Cys Val	
40	75 80 85 90	
	ATG AAG ATG TTC GCC CTG CGA CAG TAC TAC TTC ACC AAC GGC TGG AAC	878
	Met Lys Met Phe Ala Leu Arg Gln Tyr Tyr Phe Thr Asn Gly Trp Asn	
	95 100 105	
45	GTG TTC GAC TTC ATA GTG GTG ATC CTG TCC ATT GGG AGT CTG CTG TTT	926
	Val Phe Asp Phe Ile Val Val Ile Leu Ser Ile Gly Ser Leu Leu Phe	
	110 115 120	
50	TCT GCA ATC CTT AAG TCA CTG GAA AAC TAC TTC TCC CCG ACG CTC TTC	974
	Ser Ala Ile Leu Lys Ser Leu Glu Asn Tyr Phe Ser Pro Thr Leu Phe	
	125 130 135	
	CGG GTC ATC CGT CTG GCC AGG ATC GGC CGC ATC CTC AGG CTG ATC CGA	1022
55	Arg Val Ile Arg Leu Ala Arg Ile Gly Arg Ile Leu Arg Leu Ile Arg	
	140 145 150	
	GCA GCC AAG GGG ATT CGC ACG CTG CTC TTC GCC CTC ATG ATG TCC CTG	1070
	Ala Ala Lys Gly Ile Arg Thr Leu Leu Phe Ala Leu Met Met Ser Leu	
60	155 160 165 170	
	CCC GCC CTC TTC AAC ATC GGC CTC CTC CTC TTC CTC GTC ATG TTC ATC	1118
	Pro Ala Leu Phe Asn Ile Gly Leu Leu Leu Phe Leu Val Met Phe Ile	
	175 180 185	

	TAC	TCC	ATC	TTC	GGC	ATG	GCC	AGC	TTC	GCT	AAC	GTC	GTG	GAC	GAG	GCC	1166
	Tyr	Ser	Ile	Phe	Gly	Met	Ala	Ser	Phe	Ala	Asn	Val	Val	Asp	Glu	Ala	
				190					195					200			
5	GGC	ATC	GAC	GAC	ATG	TTC	AAC	TTC	AAG	ACC	TTT	GGC	AAC	AGC	ATG	CTG	1214
	Gly	Ile	Asp	Asp	Met	Phe	Asn	Phe	Lys	Thr	Phe	Gly	Asn	Ser	Met	Leu	
			205					210					215				
10	TGC	CTG	TTC	CAG	ATC	ACC	ACC	TCG	GCC	GGC	TGG	GAC	GGC	CTC	CTC	AGC	1262
	Cys	Leu	Phe	Gln	Ile	Thr	Thr	Ser	Ala	Gly	Trp	Asp	Gly	Leu	Leu	Ser	
		220					225					230					
15	CCC	ATC	CTC	AAC	ACG	GGG	CCT	CCC	TAC	TGC	GAC	CCC	AAC	CTG	CCC	AAC	1310
	Pro	Ile	Leu	Asn	Thr		Pro	Pro	Tyr	Cys	Asp	Pro	Asn	Leu	Pro	Asn	
	235					240					245					250	
20	AGC	AAC	GGC	TCC	CGG	GGG	AAC	TGC	GGG	AGC	CCG	GCG	GTG	GGC	ATC	ATC	1358
	Ser	Asn	Gly	Ser	Arg	Gly	Asn	Cys	Gly	Ser	Pro	Ala	Val	Gly	Ile	Ile	
					255					260					265		
25	TTC	TTC	ACC	ACC	TAC	ATC	ATC	ATC	TCC	TTC	CTC	ATC	GTG	GTC	AAC	ATG	1406
	Phe	Phe	Thr	Thr	Tyr	Ile	Ile	Ile	Ser	Phe	Leu	Ile	Val	Val	Asn	Met	
				270					275					280			
30	TAC	ATC	GCA	GTG	ATT	CTG	GAG	AAC	TTC	AAC	GTA	GCC	ACC	GAG	GAG	AGC	1454
	Tyr	Ile	Ala	Val	Ile	Leu	Glu	Asn	Phe	Asn	Val	Ala	Thr	Glu	Glu	Ser	
			285					290					295				
35	ACG	GAG	CCC	CTG	AGC	GAG	GAC	GAC	TTC	GAC	ATG	TTC	TAT	GAG	ACC	TGG	1502
	Thr	Glu	Pro	Leu	Ser	Glu	Asp	Asp	Phe	Asp	Met	Phe	Tyr	Glu	Thr	Trp	
		300					305					310					
40	GAG	AAG	TTC	GAC	CCG	GAG	GCC	ACC	CAG	TTC	ATT	GCC	TTT	TCT	GCC	CTC	1550
	Glu	Lys	Phe	Asp	Pro	Glu	Ala	Thr	Gln	Phe	Ile	Ala	Phe	Ser	Ala	Leu	
	315					320					325					330	
45	TCA	GAC	TTC	GCG	GAC	ACG	CTC	TCC	GGC	CCT	CTT	AGA	ATC	CCC	AAA	CCC	1598
	Ser	Asp	Phe	Ala	Asp	Thr	Leu	Ser	Gly	Pro	Leu	Arg	Ile	Pro	Lys	Pro	
					335					340					345		
50	AAC	CAG	AAT	ATA	TTA	ATC	CAG	ATG	GAC	CTG	CCG	TTG	GTC	CCC	GGG	GAT	1646
	Asn	Gln	Asn	Ile	Leu	Ile	Gln	Met	Asp	Leu	Pro	Leu	Val	Pro	Gly	Asp	
				350					355					360			
55	AAG	ATC	CAC	TGT	CTG	GAC	ATC	CTT	TTT	GCC	TTC	ACA	AAG	AAC	GTC	TTG	1694
	Lys	Ile	His	Cys	Leu	Asp	Ile	Leu	Phe	Ala	Phe	Thr	Lys	Asn	Val	Leu	
			365					370					375				
60	GGA	GAA	TCC	GGG	GAG	TTG	GAC	TCC	CTG	AAG	ACC	AAT	ATG	GAA	GAG	AAG	1742
	Gly	Glu	Ser	Gly	Glu	Leu	Asp	Ser	Leu	Lys	Thr	Asn	Met	Glu	Glu	Lys	
		380					385					390					
65	TTT	ATG	GCG	ACC	AAT	CTC	TCC	AAA	GCA	TCC	TAT	GAA	CCA	ATA	GCC	ACC	1790
	Phe	Met	Ala	Thr	Asn	Leu	Ser	Lys	Ala	Ser	Tyr	Glu	Pro	Ile	Ala	Thr	
	395					400					405					410	
70	ACC	CTC	CGG	TGG	AAG	CAG	GAA	GAC	CTC	TCA	GCC	ACA	GTC	ATT	CAA	AAG	1838
	Thr	Leu	Arg	Trp	Lys	Gln	Glu	Asp	Leu	Ser	Ala	Thr	Val	Ile	Gln	Lys	
				415						420					425		
75	GCC	TAC	CGG	AGC	TAC	ATG	CTG	CAC	CGC	TCC	TTG	ACA	CTC	TCC	AAC	ACC	1886
	Ala	Tyr	Arg	Ser	Tyr	Met	Leu	His	Arg	Ser	Leu	Thr	Leu	Ser	Asn	Thr	
				430					435						440		

-67-

CTG CAT GTG CCC AGG GCT GAG GAG GAT GGC GTG TCA CTT CCC GGG GAA 1934
 Leu His Val Pro Arg Ala Glu Glu Asp Gly Val Ser Leu Pro Gly Glu
 445 450 455
 5 GGC TAC ATT ACA TTC ATG GCA AAC AGT GGA CTC CCG GAC AAA TCA GAA 1982
 Gly Tyr Ile Thr Phe Met Ala Asn Ser Gly Leu Pro Asp Lys Ser Glu
 460 465 470
 10 ACT GCC TCT GCT ACG TCT TTC CCG CCA TCC TAT GAC AGT GTC ACC AGG 2030
 Thr Ala Ser Ala Thr Ser Phe Pro Pro Ser Tyr Asp Ser Val Thr Arg
 475 480 485 490
 15 GGC CTG AGT GAC CGG GCC AAC ATT AAC CCA TCT AGC TCA ATG CAA AAT 2078
 Gly Leu Ser Asp Arg Ala Asn Ile Asn Pro Ser Ser Ser Met Gln Asn
 495 500 505
 GAA GAT GAG GTC GCT GCT AAG GAA GGA AAC AGC CCT GGA CCT CAG TGAAGGCACT
 2133
 20 Glu Asp Glu Val Ala Ala Lys Glu Gly Asn Ser Pro Gly Pro Gln
 510 515 520
 25 CAGGCATGCA CAGGGCAGGT TCCAATGTCT TTCTCTGCTG TACTAACTCC TTCCCTCTGG 2193
 AGGTGGCACC AACCTCCAGC CTCCACCAAT GCATGTCCTT GGTCATGGTG TCAGAACTGA 2253
 ATGGGGACAT CCTTGAGAAA GCGGGGACCC CAATAGGAAT CAAAAGCCAA GGATACTCCT 2313
 30 CCATTCTGAC GTCCCTTCCG AGTTCCCAGA AGATGTCATT GCTCCCTTCT GTTTGTGACC 2373
 AGAGACGTGA TTCACCAACT TCTCGGAGCC AGAGACACAT AGCAAAGACT TTTCTGCTGG 2433
 35 TGTCGGGCAG TCTTAGAGAA GTCACGTAGG GGTTGGTACT GAGAATTAGG GTTTGCATGA 2493
 CTGCATGCTC ACAGCTGCCG GACAATACCT GTGAGTCGGC CATTAAAATT AATATTTTAA 2553
 AAGTTAAAAA AAAAAAAAAA 2573

40 (2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 521 amino acids
 45 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Thr Glu Glu Gln Lys Lys Tyr Tyr Asn Ala Met Lys Lys Leu Gly
 1 5 10 15
 55 Ser Lys Lys Pro Gln Lys Pro Ile Pro Arg Pro Leu Asn Lys Tyr Gln
 20 25 30
 Gly Phe Val Phe Asp Ile Val Thr Arg Gln Ala Phe Asp Ile Ile Ile
 35 40 45
 60 Met Val Leu Ile Cys Leu Asn Met Ile Thr Met Met Val Glu Thr Asp
 50 55 60

-68-

Glu Gln Gly Glu Glu Lys Thr Lys Val Leu Gly Arg Ile Asn Gln Phe
 65 70 75 80
 5 Phe Val Ala Val Phe Thr Gly Glu Cys Val Met Lys Met Phe Ala Leu
 85 90 95
 Arg Gln Tyr Tyr Phe Thr Asn Gly Trp Asn Val Phe Asp Phe Ile Val
 100 105 110
 10 Val Ile Leu Ser Ile Gly Ser Leu Leu Phe Ser Ala Ile Leu Lys Ser
 115 120 125
 Leu Glu Asn Tyr Phe Ser Pro Thr Leu Phe Arg Val Ile Arg Leu Ala
 130 135 140
 15 Arg Ile Gly Arg Ile Leu Arg Leu Ile Arg Ala Ala Lys Gly Ile Arg
 145 150 155 160
 Thr Leu Leu Phe Ala Leu Met Met Ser Leu Pro Ala Leu Phe Asn Ile
 165 170 175
 Gly Leu Leu Leu Phe Leu Val Met Phe Ile Tyr Ser Ile Phe Gly Met
 180 185 190
 25 Ala Ser Phe Ala Asn Val Val Asp Glu Ala Gly Ile Asp Asp Met Phe
 195 200 205
 Asn Phe Lys Thr Phe Gly Asn Ser Met Leu Cys Leu Phe Gln Ile Thr
 210 215 220
 30 Thr Ser Ala Gly Trp Asp Gly Leu Leu Ser Pro Ile Leu Asn Thr Gly
 225 230 235 240
 Pro Pro Tyr Cys Asp Pro Asn Leu Pro Asn Ser Asn Gly Ser Arg Gly
 245 250 255
 Asn Cys Gly Ser Pro Ala Val Gly Ile Ile Phe Phe Thr Thr Tyr Ile
 260 265 270
 40 Ile Ile Ser Phe Leu Ile Val Val Asn Met Tyr Ile Ala Val Ile Leu
 275 280 285
 Glu Asn Phe Asn Val Ala Thr Glu Glu Ser Thr Glu Pro Leu Ser Glu
 290 295 300
 45 Asp Asp Phe Asp Met Phe Tyr Glu Thr Trp Glu Lys Phe Asp Pro Glu
 305 310 315 320
 Ala Thr Gln Phe Ile Ala Phe Ser Ala Leu Ser Asp Phe Ala Asp Thr
 325 330 335
 50 Leu Ser Gly Pro Leu Arg Ile Pro Lys Pro Asn Gln Asn Ile Leu Ile
 340 345 350
 55 Gln Met Asp Leu Pro Leu Val Pro Gly Asp Lys Ile His Cys Leu Asp
 355 360 365
 Ile Leu Phe Ala Phe Thr Lys Asn Val Leu Gly Glu Ser Gly Glu Leu
 370 375 380
 60 Asp Ser Leu Lys Thr Asn Met Glu Glu Lys Phe Met Ala Thr Asn Leu
 385 390 395 400
 Ser Lys Ala Ser Tyr Glu Pro Ile Ala Thr Thr Leu Arg Trp Lys Gln

-69-

405 410 415
 Glu Asp Leu Ser Ala Thr Val Ile Gln Lys Ala Tyr Arg Ser Tyr Met
 420 425 430
 5 Leu His Arg Ser Leu Thr Leu Ser Asn Thr Leu His Val Pro Arg Ala
 435 440 445
 10 Glu Glu Asp Gly Val Ser Leu Pro Gly Glu Gly Tyr Ile Thr Phe Met
 450 455 460
 Ala Asn Ser Gly Leu Pro Asp Lys Ser Glu Thr Ala Ser Ala Thr Ser
 465 470 475 480
 15 Phe Pro Pro Ser Tyr Asp Ser Val Thr Arg Gly Leu Ser Asp Arg Ala
 485 490 495
 Asn Ile Asn Pro Ser Ser Ser Met Gln Asn Glu Asp Glu Val Ala Ala
 500 505 510
 20 Lys Glu Gly Asn Ser Pro Gly Pro Gln
 515 520

25 (2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 7052 base pairs
 (B) TYPE: nucleic acid
 30 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

35

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 204..6602

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TAGCTTGCTT CTGCTAATGC TACCCCAGGC CTTTAGACAG AGAACAGATG GCAGATGGAG 60
 45 TTTCTTATTG CCATGCGCAA ACGCTGAGCC CACCTCATGA TCCCGGACCC CATGGTTTTC 120
 AGTAGACAAC CTGGGCTAAG AAGAGATCTC CGACCTTATA GAGCAGCAAA GAGTGTAAT 180
 TCTTCCCCAA GAAGAATGAG AAG ATG GAG CTC CCC TTT GCG TCC GTG GGA 230
 50 Met Glu Leu Pro Phe Ala Ser Val Gly
 1 5
 ACT ACC AAT TTC AGA CGG TTC ACT CCA GAG TCA CTG GCA GAG ATC GAG 278
 55 Thr Thr Asn Phe Arg Arg Phe Thr Pro Glu Ser Leu Ala Glu Ile Glu
 10 15 20 25
 AAG CAG ATT GCT GCT CAC CGG GCA GCC AAG AAG GCC AGA ACC AAG CAC 326
 Lys Gln Ile Ala Ala His Arg Ala Ala Lys Lys Ala Arg Thr Lys His
 30 35 40
 60 AGA GGA CAG GAG GAC AAG GGC GAG AAG CCC AGG CCT CAG CTG GAC TTG 374
 Arg Gly Gln Glu Asp Lys Gly Glu Lys Pro Arg Pro Gln Leu Asp Leu
 45 50 55

-70-

	AAA	GAC	TGT	AAC	CAG	CTG	CCC	AAG	TTC	TAT	GGT	GAG	CTC	CCA	GCA	GAA	422
	Lys	Asp	Cys	Asn	Gln	Leu	Pro	Lys	Phe	Tyr	Gly	Glu	Leu	Pro	Ala	Glu	
			60					65					70				
5	CTG	GTC	GGG	GAG	CCC	CTG	GAG	GAC	CTA	GAC	CCT	TTC	TAC	AGC	ACA	CAC	470
	Leu	Val	Gly	Glu	Pro	Leu	Glu	Asp	Leu	Asp	Pro	Phe	Tyr	Ser	Thr	His	
		75					80				85						
10	CGG	ACA	TTC	ATG	GTG	TTG	AAT	AAA	AGC	AGG	ACC	ATT	TCC	AGA	TTC	AGT	518
	Arg	Thr	Phe	Met	Val	Leu	Asn	Lys	Ser	Arg	Thr	Ile	Ser	Arg	Phe	Ser	
	90					95					100					105	
15	GCC	ACT	TGG	GCC	CTG	TGG	CTC	TTC	AGT	CCC	TTC	AAC	CTG	ATC	AGA	AGA	566
	Ala	Thr	Trp	Ala	Leu	Trp	Leu	Phe	Ser	Pro	Phe	Asn	Leu	Ile	Arg	Arg	
					110					115					120		
20	ACA	GCC	ATC	AAA	GTG	TCT	GTC	CAT	TCC	TGG	TTC	TCC	ATA	TTC	ATC	ACC	614
	Thr	Ala	Ile	Lys	Val	Ser	Val	His	Ser	Trp	Phe	Ser	Ile	Phe	Ile	Thr	
				125					130					135			
25	ATC	ACT	ATT	TTG	GTC	AAC	TGC	GTG	TGC	ATG	ACC	CGA	ACT	GAT	CTT	CCA	662
	Ile	Thr	Ile	Leu	Val	Asn	Cys	Val	Cys	Met	Thr	Arg	Thr	Asp	Leu	Pro	
			140				145						150				
30	GAG	AAA	GTC	GAG	TAC	GTC	TTC	ACT	GTC	ATT	TAC	ACC	TTC	GAG	GCT	CTG	710
	Glu	Lys	Val	Glu	Tyr	Val	Phe	Thr	Val	Ile	Tyr	Thr	Phe	Glu	Ala	Leu	
		155					160					165					
35	ATT	AAG	ATA	CTG	GCA	AGA	GGG	TTT	TGT	CTA	AAT	GAG	TTC	ACT	TAT	CTT	758
	Ile	Lys	Ile	Leu	Ala	Arg	Gly	Phe	Cys	Leu	Asn	Glu	Phe	Thr	Tyr	Leu	
	170					175					180					185	
40	CGA	GAT	CCG	TGG	AAC	TGG	CTG	GAC	TTC	AGT	GTC	ATT	ACC	TTG	GCG	TAT	806
	Arg	Asp	Pro	Trp	Asn	Trp	Leu	Asp	Phe	Ser	Val	Ile	Thr	Leu	Ala	Tyr	
					190					195					200		
45	GTG	GGT	GCA	GCG	ATA	GAC	CTC	CGA	GGA	ATC	TCA	GGC	CTG	CGG	ACA	TTC	854
	Val	Gly	Ala	Ala	Ile	Asp	Leu	Arg	Gly	Ile	Ser	Gly	Leu	Arg	Thr	Phe	
				205					210					215			
50	CGA	GTT	CTC	AGA	GCC	CTG	AAA	ACT	GTT	TCT	GTG	ATC	CCA	GGA	CTG	AAG	902
	Arg	Val	Leu	Arg	Ala	Leu	Lys	Thr	Val	Ser	Val	Ile	Pro	Gly	Leu	Lys	
			220				225						230				
55	GTC	ATC	GTG	GGA	GCC	CTG	ATC	CAC	TCA	GTG	AGG	AAG	CTG	GCC	GAC	GTG	950
	Val	Ile	Val	Gly	Ala	Leu	Ile	His	Ser	Val	Arg	Lys	Leu	Ala	Asp	Val	
		235					240					245					
60	ACT	ATC	CTC	ACA	GTG	TTC	TGC	CTG	AGC	GTC	TTC	GCC	TTG	GTG	GGC	CTG	998
	Thr	Ile	Leu	Thr	Val	Phe	Cys	Leu	Ser	Val	Phe	Ala	Leu	Val	Gly	Leu	
	250					255					260					265	
65	CAG	CTC	TTT	AAG	GGG	AAC	CTT	AAG	AAC	AAA	TGC	ATC	AGG	AAC	GGA	ACA	1046
	Gln	Leu	Phe	Lys	Gly	Asn	Leu	Lys	Asn	Lys	Cys	Ile	Arg	Asn	Gly	Thr	
				270						275					280		
70	GAT	CCC	CAC	AAG	GCT	GAC	AAC	CTC	TCA	TCT	GAA	ATG	GCA	GAA	TAC	ATC	1094
	Asp	Pro	His	Lys	Ala	Asp	Asn	Leu	Ser	Ser	Glu	Met	Ala	Glu	Tyr	Ile	
				285					290					295			
75	TTC	ATC	AAG	CCT	GGT	ACT	ACG	GAT	CCC	TTC	CTG	TGC	GGC	AAT	GGG	TCT	1142
	Phe	Ile	Lys	Pro	Gly	Thr	Thr	Asp	Pro	Leu	Leu	Cys	Gly	Asn	Gly	Ser	
			300					305					310				

-71-

	GAT	GCT	GGT	CAC	TGC	CCT	GGA	GGC	TAT	GTC	TGC	CTG	AAA	ACT	CCT	GAC	1190
	Asp	Ala	Gly	His	Cys	Pro	Gly	Gly	Tyr	Val	Cys	Leu	Lys	Thr	Pro	Asp	
	315						320					325					
5	AAC	CCG	GAT	TTT	AAC	TAC	ACC	AGC	TTT	GAT	TCC	TTT	GCG	TGG	GCA	TTC	1238
	Asn	Pro	Asp	Phe	Asn	Tyr	Thr	Ser	Phe	Asp	Ser	Phe	Ala	Trp	Ala	Phe	
	330					335					340					345	
10	CTC	TCA	CTG	TTC	CGC	CTC	ATG	ACG	CAG	GAC	TCC	TGG	GAG	CGC	CTG	TAC	1286
	Leu	Ser	Leu	Phe	Arg	Leu	Met	Thr	Gln	Asp	Ser	Trp	Glu	Arg	Leu	Tyr	
					350					355					360		
15	CAG	CAG	ACA	CTC	CGG	GCT	TCT	GGG	AAA	ATG	TAC	ATG	GTC	TTT	TTC	GTG	1334
	Gln	Gln	Thr	Leu	Arg	Ala	Ser	Gly	Lys	Met	Tyr	Met	Val	Phe	Phe	Val	
				365					370					375			
20	CTG	GTT	ATT	TTC	CTT	GGA	TCG	TTC	TAC	CTG	GTC	AAT	TTG	ATC	TTG	GCC	1382
	Leu	Val	Ile	Phe	Leu	Gly	Ser	Phe	Tyr	Leu	Val	Asn	Leu	Ile	Leu	Ala	
			380					385					390				
25	GTG	GTC	ACC	ATG	GCG	TAT	GAA	GAG	CAG	AGC	CAG	GCA	ACA	ATT	GCA	GAA	1430
	Val	Val	Thr	Met	Ala	Tyr	Glu	Glu	Gln	Ser	Gln	Ala	Thr	Ile	Ala	Glu	
		395					400					405					
30	ATC	GAA	GCC	AAG	GAA	AAA	AAG	TTC	CAG	GAA	GCC	CTT	GAG	GTG	CTG	CAG	1478
	Ile	Glu	Ala	Lys	Glu	Lys	Lys	Phe	Gln	Glu	Ala	Leu	Glu	Val	Leu	Gln	
	410					415					420					425	
35	AAG	GAA	CAG	GAG	GTG	CTG	GCA	GCC	CTG	GGG	ATT	GAC	ACG	ACC	TCG	CTC	1526
	Lys	Glu	Gln	Glu	Val	Leu	Ala	Ala	Leu	Gly	Ile	Asp	Thr	Thr	Ser	Leu	
				430						435					440		
40	CAG	TCC	CAC	AGT	GGA	TCA	CCC	TTA	GCC	TCC	AAA	AAC	GCC	AAT	GAG	AGA	1574
	Gln	Ser	His	Ser	Gly	Ser	Pro	Leu	Ala	Ser	Lys	Asn	Ala	Asn	Glu	Arg	
				445					450					455			
45	AGA	CCC	AGG	GTG	AAA	TCA	AGG	GTG	TCA	GAG	GGC	TCC	ACG	GAT	GAC	AAC	1622
	Arg	Pro	Arg	Val	Lys	Ser	Arg	Val	Ser	Glu	Gly	Ser	Thr	Asp	Asp	Asn	
			460					465					470				
50	AGG	TCA	CCC	CAA	TCT	GAC	CCT	TAC	AAC	CAG	CGC	AGG	ATG	TCT	TTC	CTA	1670
	Arg	Ser	Pro	Gln	Ser	Asp	Pro	Tyr	Asn	Gln	Arg	Arg	Met	Ser	Phe	Leu	
		475					480					485					
55	GGC	CTG	TCT	TCA	GGA	AGA	CGC	AGG	GCT	AGC	CAC	GGC	AGT	GTG	TTC	CAC	1718
	Gly	Leu	Ser	Ser	Gly	Arg	Arg	Arg	Ala	Ser	His	Gly	Ser	Val	Phe	His	
	490				495						500					505	
60	TTC	CGA	GCG	CCC	AGC	CAA	GAC	ATC	TCA	TTT	CCT	GAC	GGG	ATC	ACC	CCT	1766
	Phe	Arg	Ala	Pro	Ser	Gln	Asp	Ile	Ser	Phe	Pro	Asp	Gly	Ile	Thr	Pro	
				510						515					520		
65	GAT	GAT	GGG	GTC	TTT	CAC	GGA	GAC	CAG	GAA	AGC	CGT	CGA	GGT	TCC	ATA	1814
	Asp	Asp	Gly	Val	Phe	His	Gly	Asp	Gln	Glu	Ser	Arg	Arg	Gly	Ser	Ile	
				525					530					535			
70	TTG	CTG	GGC	AGG	GGT	GCT	GGG	CAG	ACA	GGT	CCA	CTC	CCC	AGG	AGC	CCA	1862
	Leu	Leu	Gly	Arg	Gly	Ala	Gly	Gln	Thr	Gly	Pro	Leu	Pro	Arg	Ser	Pro	
			540				545						550				
75	CTG	CCT	CAG	TCC	CCC	AAC	CCT	GGC	CGT	AGA	CAT	GGA	GAA	GAG	GGA	CAG	1910
	Leu	Pro	Gln	Ser	Pro	Asn	Pro	Gly	Arg	Arg	His	Gly	Glu	Glu	Gly	Gln	
		555					560					565					

	CTC	GGA	GTG	CCC	ACT	GGT	GAG	CTT	ACC	GCT	GGA	GCG	CCT	GAA	GGC	CCG	1958
	Leu	Gly	Val	Pro	Thr	Gly	Glu	Leu	Thr	Ala	Gly	Ala	Pro	Glu	Gly	Pro	
	570					575					580					585	
5	GCA	CTC	GAC	ACT	ACA	GGG	CAG	AAG	AGC	TTC	CTG	TCT	GCG	GGC	TAC	TTG	2006
	Ala	Leu	Asp	Thr	Thr	Gly	Gln	Lys	Ser	Phe	Leu	Ser	Ala	Gly	Tyr	Leu	
					590					595					600		
10	AAC	GAA	CCT	TTC	CGA	GCA	CAG	AGG	GCC	ATG	AGC	GTT	GTC	AGT	ATC	ATG	2054
	Asn	Glu	Pro	Phe	Arg	Ala	Gln	Arg	Ala	Met	Ser	Val	Val	Ser	Ile	Met	
				605					610					615			
15	ACT	TCT	GTC	ATT	GAG	GAG	CTT	GAA	GAG	TCT	AAG	CTG	AAG	TGC	CCA	CCC	2102
	Thr	Ser	Val	Ile	Glu	Glu	Leu	Glu	Glu	Ser	Lys	Leu	Lys	Cys	Pro	Pro	
			620					625					630				
20	TGC	TTG	ATC	AGC	TTC	GCT	CAG	AAG	TAT	CTG	ATC	TGG	GAG	TGC	TGC	CCC	2150
	Cys	Leu	Ile	Ser	Phe	Ala	Gln	Lys	Tyr	Leu	Ile	Trp	Glu	Cys	Cys	Pro	
		635					640					645					
25	AAG	TGG	AGG	AAG	TTC	AAG	ATG	GCG	CTG	TTC	GAG	CTG	GTG	ACT	GAC	CCC	2198
	Lys	Trp	Arg	Lys	Phe	Lys	Met	Ala	Leu	Phe	Glu	Leu	Val	Thr	Asp	Pro	
		650				655					660					665	
30	TTC	GCA	GAG	CTT	ACC	ATC	ACC	CTC	TGC	ATC	GTG	GTG	AAC	ACC	GTC	TTC	2246
	Phe	Ala	Glu	Leu	Thr	Ile	Thr	Leu	Cys	Ile	Val	Val	Asn	Thr	Val	Phe	
					670					675					680		
35	ATG	GCC	ATG	GAG	CAC	TAC	CCC	ATG	ACC	GAT	GCC	TTC	GAT	GCC	ATG	CTT	2294
	Met	Ala	Met	Glu	His	Tyr	Pro	Met	Thr	Asp	Ala	Phe	Asp	Ala	Met	Leu	
				685					690					695			
40	CAA	GCC	GGC	AAC	ATT	GTC	TTC	ACC	GTG	TTT	TTC	ACA	ATG	GAG	ATG	GCC	2342
	Gln	Ala	Gly	Asn	Ile	Val	Phe	Thr	Val	Phe	Phe	Thr	Met	Glu	Met	Ala	
			700					705					710				
45	TTC	AAG	ATC	ATT	GCC	TTC	GAC	CCC	TAC	TAT	TAC	TTC	CAG	AAG	AAG	TGG	2390
	Phe	Lys	Ile	Ile	Ala	Phe	Asp	Pro	Tyr	Tyr	Tyr	Phe	Gln	Lys	Lys	Trp	
		715					720					725					
50	AAT	ATC	TTC	GAC	TGT	GTC	ATC	GTC	ACC	GTG	AGC	CTT	CTG	GAG	CTG	AGT	2438
	Asn	Ile	Phe	Asp	Cys	Val	Ile	Val	Thr	Val	Ser	Leu	Leu	Glu	Leu	Ser	
		730				735					740					745	
55	GCA	TCC	AAG	AAG	GGC	AGC	CTG	TCT	GTG	CTC	CGT	TCC	TTA	CGC	TTG	GCA	2486
	Ala	Ser	Lys	Lys	Gly	Ser	Leu	Ser	Val	Leu	Arg	Ser	Leu	Arg	Leu	Ala	
					750					755					760		
60	CTC	GAC	ACT	ACA	GGG	CAG	AAG	AGC	TTC	CTG	TCT	GCG	GGC	TAC	TTG	AAC	2534
	Leu	Asp	Thr	Thr	Gly	Gln	Lys	Ser	Phe	Leu	Ser	Ala	Gly	Tyr	Leu	Asn	
				765					770					775			
65	GAA	CCT	TTC	CGA	GCA	CAG	AGG	GCC	ATG	AGC	GTT	GTC	AGT	ATC	ATG	ACT	2582
	Glu	Pro	Phe	Arg	Ala	Gln	Arg	Ala	Met	Ser	Val	Val	Ser	Ile	Met	Thr	
			780					785					790				
70	TCT	GTC	ATT	GAG	GAG	CTT	GAA	GAG	TCT	AAG	CTG	AAG	TGC	CCA	CCC	TGC	2630
	Ser	Val	Ile	Glu	Glu	Leu	Glu	Glu	Ser	Lys	Leu	Lys	Cys	Pro	Pro	Cys	
		795					800					805					
75	TTG	ATC	AGC	TTC	GCT	CAG	AAG	TAT	CTG	ATC	TGG	GAG	TGC	TGC	CCC	AAG	2678
	Leu	Ile	Ser	Phe	Ala	Gln	Lys	Tyr	Leu	Ile	Trp	Glu	Cys	Cys	Pro	Lys	
		810				815					820					825	

	TGG	AGG	AAG	TTC	AAG	ATG	GCG	CTG	TTC	GAG	CTG	GTG	ACT	GAC	CCC	TTC	2726
	Trp	Arg	Lys	Phe	Lys	Met	Ala	Leu	Phe	Glu	Leu	Val	Thr	Asp	Pro	Phe	
					830					835					840		
5	GCA	GAG	CTT	ACC	ATC	ACC	CTC	TGC	ATC	GTG	GTG	AAC	ACC	GTC	TTC	ATG	2774
	Ala	Glu	Leu	Thr	Ile	Thr	Leu	Cys	Ile	Val	Val	Asn	Thr	Val	Phe	Met	
				845					850					855			
10	GCC	ATG	GAG	CAC	TAC	CCC	ATG	ACC	GAT	GCC	TTC	GAT	GCC	ATG	CTT	CAA	2822
	Ala	Met	Glu	His	Tyr	Pro	Met	Thr	Asp	Ala	Phe	Asp	Ala	Met	Leu	Gln	
			860					865					870				
15	GCC	GGC	AAC	ATT	GTC	TTC	ACC	GTG	TTT	TTC	ACA	ATG	GAG	ATG	GCC	TTC	2870
	Ala	Gly	Asn	Ile	Val	Phe	Thr	Val	Phe	Phe	Thr	Met	Glu	Met	Ala	Phe	
		875						880					885				
20	AAG	ATC	ATT	GCC	TTC	GAC	CCC	TAC	TAT	TAC	TTC	CAG	AAG	AAG	TGG	AAT	2918
	Lys	Ile	Ile	Ala	Phe	Asp	Pro	Tyr	Tyr	Tyr	Phe	Gln	Lys	Lys	Trp	Asn	
	890					895					900					905	
25	ATC	TTC	GAC	TGT	GTC	ATC	GTC	ACC	GTG	AGC	CTT	CTG	GAG	CTG	AGT	GCA	2966
	Ile	Phe	Asp	Cys	Val	Ile	Val	Thr	Val	Ser	Leu	Leu	Glu	Leu	Ser	Ala	
					910					915					920		
30	TCC	AAG	AAG	GGC	AGC	CTG	TCT	GTG	CTC	CGT	TCC	TTA	CGC	TTG	CTG	CGG	3014
	Ser	Lys	Lys	Gly	Ser	Leu	Ser	Val	Leu	Arg	Ser	Leu	Arg	Leu	Leu	Arg	
				925					930					935			
35	GTC	TTC	AAG	CTG	GCC	AAG	TCC	TGG	CCC	ACC	CTG	AAC	ACC	CTC	ATC	AAG	3062
	Val	Phe	Lys	Leu	Ala	Lys	Ser	Trp	Pro	Thr	Leu	Asn	Thr	Leu	Ile	Lys	
			940					945					950				
40	ATC	ATC	GGG	AAC	TCA	GTG	GGG	GCC	CTG	GGC	AAC	CTG	ACC	TTT	ATC	CTG	3110
	Ile	Ile	Gly	Asn	Ser	Val	Gly	Ala	Leu	Gly	Asn	Leu	Thr	Phe	Ile	Leu	
		955					960					965					
45	GCC	ATC	ATC	GTC	TTC	ATC	TTC	GCC	CTG	GTC	GGA	AAG	CAG	CTT	CTC	TCA	3158
	Ala	Ile	Ile	Val	Phe	Ile	Phe	Ala	Leu	Val	Gly	Lys	Gln	Leu	Leu	Ser	
	970					975					980					985	
50	GAG	GAC	TAC	GGG	TGC	CGC	AAG	GAC	GGC	GTC	TCC	GTG	TGG	AAC	GGC	GAG	3206
	Glu	Asp	Tyr	Gly	Cys	Arg	Lys	Asp	Gly	Val	Ser	Val	Trp	Asn	Gly	Glu	
				990					995						1000		
55	AAG	CTC	CGC	TGG	CAC	ATG	TGT	GAC	TTC	TTC	CAT	TCC	TTC	CTG	GTC	GTC	3254
	Lys	Leu	Arg	Trp	His	Met	Cys	Asp	Phe	Phe	His	Ser	Phe	Leu	Val	Val	
				1005					1010					1015			
60	TTC	CGA	ATC	CTC	TGC	GGG	GAG	TGG	ATC	GAG	AAC	ATG	TGG	GTC	TGC	ATG	3302
	Phe	Arg	Ile	Leu	Cys	Gly	Glu	Trp	Ile	Glu	Asn	Met	Trp	Val	Cys	Met	
			1020					1025					1030				
65	GAG	GTC	AGC	CAG	AAA	TCC	ATC	TGC	CTC	ATC	CTC	TTC	TTG	ACT	GTG	ATG	3350
	Glu	Val	Ser	Gln	Lys	Ser	Ile	Cys	Leu	Ile	Leu	Phe	Leu	Thr	Val	Met	
		1035					1040					1045					
70	GTG	CTG	GGC	AAC	CTA	GTG	GTG	CTC	AAC	CTT	TTC	ATC	GCT	TTA	CTG	CTG	3398
	Val	Leu	Gly	Asn	Leu	Val	Val	Leu	Asn	Leu	Phe	Ile	Ala	Leu	Leu	Leu	
	1050					1055					1060					1065	
75	AAC	TCC	TTC	AGC	GCG	GAC	AAC	CTC	ACG	GCT	CCA	GAG	GAT	GAC	GGG	GAG	3446
	Asn	Ser	Phe	Ser	Ala	Asp	Asn	Leu	Thr	Ala	Pro	Glu	Asp	Asp	Gly	Glu	
					1070					1075					1080		

	GTG AAC AAC TTG CAG TTA GCA CTG GCC AGG ATC CAG GTA CTT GGC CAT	3494
	Val Asn Asn Leu Gln Leu Ala Leu Ala Arg Ile Gln Val Leu Gly His	
	1085 1090 1095	
5	CGG GCC AGC AGG GCC ATC GCC AGT TAC ATC AGC AGC CAC TGC CGA TTC	3542
	Arg Ala Ser Arg Ala Ile Ala Ser Tyr Ile Ser Ser His Cys Arg Phe	
	1100 1105 1110	
10	CGC TGG CCC AAG GTG GAG ACC CAG CTG GGC ATG AAG CCC CCA CTC ACC	3590
	Arg Trp Pro Lys Val Glu Thr Gln Leu Gly Met Lys Pro Pro Leu Thr	
	1115 1120 1125	
15	AGC TCA GAG GCC AAG AAC CAC ATT GCC ACT GAT GCT GTC AGT GCT GCA	3638
	Ser Ser Glu Ala Lys Asn His Ile Ala Thr Asp Ala Val Ser Ala Ala	
	1130 1135 1140 1145	
20	GTG GGG AAC CTG ACA AAG CCA GCT CTC AGT AGC CCC AAG GAG AAT CAC	3686
	Val Gly Asn Leu Thr Lys Pro Ala Leu Ser Ser Pro Lys Glu Asn His	
	1150 1155 1160	
25	GGG GAC TTC ATC ACT GAT CCC AAC GTG TGG GTC TCT GTG CCC ATT GCT	3734
	Gly Asp Phe Ile Thr Asp Pro Asn Val Trp Val Ser Val Pro Ile Ala	
	1165 1170 1175	
30	GAG GGG GAA TCT GAC CTC GAC GAG CTC GAG GAA GAT ATG GAG CAG GCT	3782
	Glu Gly Glu Ser Asp Leu Asp Glu Leu Glu Glu Asp Met Glu Gln Ala	
	1180 1185 1190	
35	TCG CAG AGC TCC TGG CAG GAA GAG GAC CCC AAG GGA CAG CAG GAG CAG	3830
	Ser Gln Ser Ser Trp Gln Glu Glu Asp Pro Lys Gly Gln Gln Glu Gln	
	1195 1200 1205	
40	TTG CCA CAA GTC CAA AAG TGT GAA AAC CAC CAG GCA GCC AGA AGC CCA	3878
	Leu Pro Gln Val Gln Lys Cys Glu Asn His Gln Ala Ala Arg Ser Pro	
	1210 1215 1220 1225	
45	GCC TCC ATG ATG TCC TCT GAG GAC CTG GCT CCA TAC CTG GGT GAG AGC	3926
	Ala Ser Met Met Ser Ser Glu Asp Leu Ala Pro Tyr Leu Gly Glu Ser	
	1230 1235 1240	
50	TGG AAG AGG AAG GAT AGC CCT CAG GTC CCT GCC GAG GGA GTG GAT GAC	3974
	Trp Lys Arg Lys Asp Ser Pro Gln Val Pro Ala Glu Gly Val Asp Asp	
	1245 1250 1255	
55	ACG AGC TCC TCT GAG GGC AGC ACG GTG GAC TGC CCG GAC CCA GAG GAA	4022
	Thr Ser Ser Ser Glu Gly Ser Thr Val Asp Cys Pro Asp Pro Glu Glu	
	1260 1265 1270	
60	ATC CTG AGG AAG ATC CCC GAG CTG GCA GAT GAC CTG GAC GAG CCC GAT	4070
	Ile Leu Arg Lys Ile Pro Glu Leu Ala Asp Asp Leu Asp Glu Pro Asp	
	1275 1280 1285	
65	GAC TGT TTC ACA GAA GGC TGC ACT CGC CGC TGT CCC TGC TGC AAC GTG	4118
	Asp Cys Phe Thr Glu Gly Cys Thr Arg Arg Cys Pro Cys Cys Asn Val	
	1290 1295 1300 1305	
70	AAT ACT AGC AAG TCT CCT TGG GCC ACA GGC TGG CAG GTG CGC AAG ACC	4166
	Asn Thr Ser Lys Ser Pro Trp Ala Thr Gly Trp Gln Val Arg Lys Thr	
	1310 1315 1320	
75	TGC TAC CGC ATC GTG GAG CAC AGC TGG TTT GAG AGT TTC ATC ATC TTC	4214
	Cys Tyr Arg Ile Val Glu His Ser Trp Phe Glu Ser Phe Ile Ile Phe	
	1325 1330 1335	

	ATG ATC CTG CTC AGC AGT GGA GCG CTG GCC TTT GAG GAT AAC TAC CTG	4262
	Met Ile Leu Leu Ser Ser Gly Ala Leu Ala Phe Glu Asp Asn Tyr Leu	
	1340 1345 1350	
5	GAA GAG AAA CCC CGA GTG AAG TCC GTG CTG GAG TAC ACT GAC CGA GTG	4310
	Glu Glu Lys Pro Arg Val Lys Ser Val Leu Glu Tyr Thr Asp Arg Val	
	1355 1360 1365	
10	TTC ACC TTC ATC TTC GTC TTT GAG ATG CTG CTC AAG TGG GTA GCC TAT	4358
	Phe Thr Phe Ile Phe Val Phe Glu Met Leu Leu Lys Trp Val Ala Tyr	
	1370 1375 1380 1385	
15	GGC TTC AAA AAG TAT TTC ACC AAT GCC TGG TGC TGG CTG GAC TTC CTC	4406
	Gly Phe Lys Lys Tyr Phe Thr Asn Ala Trp Cys Trp Leu Asp Phe Leu	
	1390 1395 1400	
20	ATT GTG AAC ATC TCC CTG ACA AGC CTC ATA GCG AAG ATC CTT GAG TAT	4454
	Ile Val Asn Ile Ser Leu Thr Ser Leu Ile Ala Lys Ile Leu Glu Tyr	
	1405 1410 1415	
25	TCC GAC GTG GCG TCC ATC AAA GCC CTT CGG ACT CTC CGT GCC CTC CGA	4502
	Ser Asp Val Ala Ser Ile Lys Ala Leu Arg Thr Leu Arg Ala Leu Arg	
	1420 1425 1430	
30	CCG CTG CGG GCT CTG TCT CGA TTC GAA GGC ATG AGG GTA GTG GTG GAT	4550
	Pro Leu Arg Ala Leu Ser Arg Phe Glu Gly Met Arg Val Val Val Asp	
	1435 1440 1445	
35	GCC CTC GTG GGC GCC ATC CCC TCC ATC ATG AAC GTC CTC CTC GTC TGC	4598
	Ala Leu Val Gly Ala Ile Pro Ser Ile Met Asn Val Leu Leu Val Cys	
	1450 1455 1460 1465	
40	CTC ATC TTC TGG CTC ATC TTC AGC ATC ATG GGC GTG AAC CTC TTC GCC	4646
	Leu Ile Phe Trp Leu Ile Phe Ser Ile Met Gly Val Asn Leu Phe Ala	
	1470 1475 1480	
45	GGG AAA TTT TCG AAG TGC GTC GAC ACC AGA AAT AAC CCA TTT TCC AAC	4694
	Gly Lys Phe Ser Lys Cys Val Asp Thr Arg Asn Asn Pro Phe Ser Asn	
	1485 1490 1495	
50	GTG AAT TCG ACG ATG GTG AAT AAC AAG TCC GAG TGT CAC AAT CAA AAC	4742
	Val Asn Ser Thr Met Val Asn Asn Lys Ser Glu Cys His Asn Gln Asn	
	1500 1505 1510	
55	AGC ACC GGC CAC TTC TTC TGG GTC AAC GTC AAA GTC AAC TTC GAC AAC	4790
	Ser Thr Gly His Phe Phe Trp Val Asn Val Lys Val Asn Phe Asp Asn	
	1515 1520 1525	
60	GTC GCT ATG GGC TAC CTC GCA CTT CTT CAG GTG GCA ACC TTC AAA GGC	4838
	Val Ala Met Gly Tyr Leu Ala Leu Leu Gln Val Ala Thr Phe Lys Gly	
	1530 1535 1540 1545	
65	TGG ATG GAC ATA ATG TAT GCA GCT GTT GAT TCC GGA GAG ATC AAC AGT	4886
	Trp Met Asp Ile Met Tyr Ala Ala Val Asp Ser Gly Glu Ile Asn Ser	
	1550 1555 1560	
70	CAG CCT AAC TGG GAG AAC AAC TTG TAC ATG TAC CTG TAC TTC GTC GTT	4934
	Gln Pro Asn Trp Glu Asn Asn Leu Tyr Met Tyr Leu Tyr Phe Val Val	
	1565 1570 1575	
75	TTC ATC ATT TTC GGT GGC TTC TTC ACG CTG AAT CTC TTT GTT GGG GTC	4982
	Phe Ile Ile Phe Gly Gly Phe Phe Thr Leu Asn Leu Phe Val Gly Val	
	1580 1585 1590	

	ATA ATC GAC AAC TTC AAC CAA CAG AAA AAA AAG CTA GGA GGC CAG GAC	5030
	Ile Ile Asp Asn Phe Asn Gln Gln Lys Lys Lys Leu Gly Gly Gln Asp	
	1595 1600 1605	
5	ATC TTC ATG ACA GAA GAG CAG AAG AAG TAC TAC AAT GCC ATG AAG AAG	5078
	Ile Phe Met Thr Glu Glu Gln Lys Lys Tyr Tyr Asn Ala Met Lys Lys	
	1610 1615 1620 1625	
10	CTG GGC TCC AAG AAA CCC CAG AAG CCC ATC CCA CGG CCC CTG AAT AAG	5126
	Leu Gly Ser Lys Lys Pro Gln Lys Pro Ile Pro Arg Pro Leu Asn Lys	
	1630 1635 1640	
15	TAC CAA GGC TTC GTG TTT GAC ATC GTG ACC AGG CAA GCC TTT GAC ATC	5174
	Tyr Gln Gly Phe Val Phe Asp Ile Val Thr Arg Gln Ala Phe Asp Ile	
	1645 1650 1655	
20	ATC ATC ATG GTT CTC ATC TGC CTC AAC ATG ATC ACC ATG ATG GTG GAG	5222
	Ile Ile Met Val Leu Ile Cys Leu Asn Met Ile Thr Met Met Val Glu	
	1660 1665 1670	
25	ACC GAC GAG CAG GGC GAG GAG AAG ACG AAG GTT CTG GGC AGA ATC AAC	5270
	Thr Asp Glu Gln Gly Glu Glu Lys Thr Lys Val Leu Gly Arg Ile Asn	
	1675 1680 1685	
30	CAG TTC TTT GTG GCC GTC TTC ACG GGC GAG TGT GTG ATG AAG ATG TTC	5318
	Gln Phe Phe Val Ala Val Phe Thr Gly Glu Cys Val Met Lys Met Phe	
	1690 1695 1700 1705	
35	GCC CTG CGA CAG TAC TAC TTC ACC AAC GGC TGG AAC GTG TTC GAC TTC	5366
	Ala Leu Arg Gln Tyr Tyr Phe Thr Asn Gly Trp Asn Val Phe Asp Phe	
	1710 1715 1720	
40	ATA GTG GTG ATC CTG TCC ATT GGG AGT CTG CTG TTT TCT GCA ATC CTT	5414
	Ile Val Val Ile Leu Ser Ile Gly Ser Leu Leu Phe Ser Ala Ile Leu	
	1725 1730 1735	
45	AAG TCA CTG GAA AAC TAC TTC TCC CCG ACG CTC TTC CGG GTC ATC CGT	5462
	Lys Ser Leu Glu Asn Tyr Phe Ser Pro Thr Leu Phe Arg Val Ile Arg	
	1740 1745 1750	
50	CTG GCC AGG ATC GGC CGC ATC CTC AGG CTG ATC CGA GCA GCC AAG GGG	5510
	Leu Ala Arg Ile Gly Arg Ile Leu Arg Leu Ile Arg Ala Ala Lys Gly	
	1755 1760 1765	
55	ATT CGC ACG CTG CTC TTC GCC CTC ATG ATG TCC CTG CCC GCC CTC TTC	5558
	Ile Arg Thr Leu Leu Phe Ala Leu Met Met Ser Leu Pro Ala Leu Phe	
	1770 1775 1780 1785	
60	AAC ATC GGC CTC CTC CTC TTC CTC GTC ATG TTC ATC TAC TCC ATC TTC	5606
	Asn Ile Gly Leu Leu Leu Phe Leu Val Met Phe Ile Tyr Ser Ile Phe	
	1790 1795 1800	
65	GGC ATG GCC AGC TTC GCT AAC GTC GTG GAC GAG GCC GGC ATC GAC GAC	5654
	Gly Met Ala Ser Phe Ala Asn Val Val Asp Glu Ala Gly Ile Asp Asp	
	1805 1810 1815	
70	ATG TTC AAC TTC AAG ACC TTT GGC AAC AGC ATG CTG TGC CTG TTC CAG	5702
	Met Phe Asn Phe Lys Thr Phe Gly Asn Ser Met Leu Cys Leu Phe Gln	
	1820 1825 1830	
75	ATC ACC ACC TCG GCC GGC TGG GAC GGC CTC CTC AGC CCC ATC CTC AAC	5750
	Ile Thr Thr Ser Ala Gly Trp Asp Gly Leu Leu Ser Pro Ile Leu Asn	
	1835 1840 1845	

-77-

	ACG GGG CCT CCC TAC TGC GAC CCC AAC CTG CCC AAC AGC AAC GGC TCC	5798
	Thr Gly Pro Pro Tyr Cys Asp Pro Asn Leu Pro Asn Ser Asn Gly Ser	
	1850 1855 1860 1865	
5	CGG GGG AAC TGC GGG AGC CCG GCG GTG GGC ATC ATC TTC TTC ACC ACC	5846
	Arg Gly Asn Cys Gly Ser Pro Ala Val Gly Ile Ile Phe Phe Thr Thr	
	1870 1875 1880	
10	TAC ATC ATC ATC TCC TTC CTC ATC GTG GTC AAC ATG TAC ATC GCA GTG	5894
	Tyr Ile Ile Ile Ser Phe Leu Ile Val Val Asn Met Tyr Ile Ala Val	
	1885 1890 1895	
15	ATT CTG GAG AAC TTC AAC GTA GCC ACC GAG GAG AGC ACG GAG CCC CTG	5942
	Ile Leu Glu Asn Phe Asn Val Ala Thr Glu Glu Ser Thr Glu Pro Leu	
	1900 1905 1910	
20	AGC GAG GAC GAC TTC GAC ATG TTC TAT GAG ACC TGG GAG AAG TTC GAC	5990
	Ser Glu Asp Asp Phe Asp Met Phe Tyr Glu Thr Trp Glu Lys Phe Asp	
	1915 1920 1925	
25	CCG GAG GCC ACC CAG TTC ATT GCC TTT TCT GCC CTC TCA GAC TTC GCG	6038
	Pro Glu Ala Thr Gln Phe Ile Ala Phe Ser Ala Leu Ser Asp Phe Ala	
	1930 1935 1940 1945	
30	GAC ACG CTC TCC GGC CCT CTT AGA ATC CCC AAA CCC AAC CAG AAT ATA	6086
	Asp Thr Leu Ser Gly Pro Leu Arg Ile Pro Lys Pro Asn Gln Asn Ile	
	1950 1955 1960	
35	TTA ATC CAG ATG GAC CTG CCG TTG GTC CCC GGG GAT AAG ATC CAC TGT	6134
	Leu Ile Gln Met Asp Leu Pro Leu Val Pro Gly Asp Lys Ile His Cys	
	1965 1970 1975	
40	CTG GAC ATC CTT TTT GCC TTC ACA AAG AAC GTC TTG GGA GAA TCC GGG	6182
	Leu Asp Ile Leu Phe Ala Phe Thr Lys Asn Val Leu Gly Glu Ser Gly	
	1980 1985 1990	
45	GAG TTG GAC TCC CTG AAG ACC AAT ATG GAA GAG AAG TTT ATG GCG ACC	6230
	Glu Leu Asp Ser Leu Lys Thr Asn Met Glu Glu Lys Phe Met Ala Thr	
	1995 2000 2005	
50	AAT CTC TCC AAA GCA TCC TAT GAA CCA ATA GCC ACC ACC CTC CGG TGG	6278
	Asn Leu Ser Lys Ala Ser Tyr Glu Pro Ile Ala Thr Thr Leu Arg Trp	
	2010 2015 2020 2025	
55	AAG CAG GAA GAC CTC TCA GCC ACA GTC ATT CAA AAG GCC TAC CGG AGC	6326
	Lys Gln Glu Asp Leu Ser Ala Thr Val Ile Gln Lys Ala Tyr Arg Ser	
	2030 2035 2040	
60	TAC ATG CTG CAC CGC TCC TTG ACA CTC TCC AAC ACC CTG CAT GTG CCC	6374
	Tyr Met Leu His Arg Ser Leu Thr Leu Ser Asn Thr Leu His Val Pro	
	2045 2050 2055	
65	AGG GCT GAG GAG GAT GGC GTG TCA CTT CCC GGG GAA GGC TAC AGT ACA	6422
	Arg Ala Glu Glu Asp Gly Val Ser Leu Pro Gly Glu Gly Tyr Ser Thr	
	2060 2065 2070	
70	TTC ATG GCA AAC AGT GGA CTC CCG GAC AAA TCA GAA ACT GCC TCT GCT	6470
	Phe Met Ala Asn Ser Gly Leu Pro Asp Lys Ser Glu Thr Ala Ser Ala	
	2075 2080 2085	
75	ACG TCT TTC CCG CCA TCC TAT GAC AGT GTC ACC AGG GGC CTG AGT GAC	6518
	Thr Ser Phe Pro Pro Ser Tyr Asp Ser Val Thr Arg Gly Leu Ser Asp	
	2090 2095 2100 2105	

-78-

CGG GCC AAC ATT AAC CCA TCT AGC TCA ATG CAA AAT GAA GAT GAG GTC 6566
 Arg Ala Asn Ile Asn Pro Ser Ser Ser Met Gln Asn Glu Asp Glu Val
 2110 2115 2120

5 GCT GCT AAG GAA GGA AAC AGC CCT GGA CCT CAG TGAAGGCACT CAGGCATGCA 6619
 Ala Ala Lys Glu Gly Asn Ser Pro Gly Pro Gln
 2125 2130

10 CAGGGCAGGT TCCAATGTCT TTCTCTGCTG TACTAACTCC TTCCCTCTGG AGGTGGCACC 6679
 AACCTCCAGC CTCCACCAAT GCATGTCACCT GGTCATGGTG TCAGAACTGA ATGGGGACAT 6739
 CCTTGAGAAA GCCCCCAGCC CAATAGGAAT CAAAAGCCAA GGATACTCCT CCATTCTGAC 6799

15 GTCCCTTCCG AGTTCCCAGA AGATGTCATT GCTCCCTTCT GTTTGTGACC AGAGACGTGA 6859
 TTCACCAACT TCTCGGAGCC AGAGACACAT AGCAAAGACT TTTCTGCTGG TGTCGGGCAG 6919
 TCTTAGAGAA GTCACGTAGG GGTTGGTACT GAGAATTAGG GTTTGCATGA CTGCATGCTC 6979

20 ACAGCTGCCG GACAATACCT GTGAGTCGGC CATTAAATT AATATTTTAA AAGTTAAAAA 7039
 AAAAAAAAAA AAA 7052

25 (2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2132 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: protein

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Glu Leu Pro Phe Ala Ser Val Gly Thr Thr Asn Phe Arg Arg Phe
 1 5 10 15

40 Thr Pro Glu Ser Leu Ala Glu Ile Glu Lys Gln Ile Ala Ala His Arg
 20 25 30

Ala Ala Lys Lys Ala Arg Thr Lys His Arg Gly Gln Glu Asp Lys Gly
 35 40 45

45 Glu Lys Pro Arg Pro Gln Leu Asp Leu Lys Asp Cys Asn Gln Leu Pro
 50 55 60

50 Lys Phe Tyr Gly Glu Leu Pro Ala Glu Leu Val Gly Glu Pro Leu Glu
 65 70 75 80

Asp Leu Asp Pro Phe Tyr Ser Thr His Arg Thr Phe Met Val Leu Asn
 85 90 95

55 Lys Ser Arg Thr Ile Ser Arg Phe Ser Ala Thr Trp Ala Leu Trp Leu
 100 105 110

Phe Ser Pro Phe Asn Leu Ile Arg Arg Thr Ala Ile Lys Val Ser Val
 115 120 125

60 His Ser Trp Phe Ser Ile Phe Ile Thr Ile Thr Ile Leu Val Asn Cys
 130 135 140

Val Cys Met Thr Arg Thr Asp Leu Pro Glu Lys Val Glu Tyr Val Phe

-79-

	145		150		155		160
	Thr Val Ile Tyr	Thr Phe Glu Ala Leu	Ile Lys Ile Leu Ala Arg Gly				
		165	170			175	
5	Phe Cys Leu Asn Glu Phe Thr Tyr	Leu Arg Asp Pro Trp Asn Trp Leu					
		180	185			190	
10	Asp Phe Ser Val Ile Thr Leu Ala Tyr Val Gly Ala Ala Ile Asp Leu						
		195	200			205	
	Arg Gly Ile Ser Gly Leu Arg Thr Phe Arg Val Leu Arg Ala Leu Lys						
		210	215			220	
15	Thr Val Ser Val Ile Pro Gly Leu Lys Val Ile Val Gly Ala Leu Ile						
		225	230			235	240
	His Ser Val Arg Lys Leu Ala Asp Val Thr Ile Leu Thr Val Phe Cys						
		245	250			255	
20	Leu Ser Val Phe Ala Leu Val Gly Leu Gln Leu Phe Lys Gly Asn Leu						
		260	265			270	
	Lys Asn Lys Cys Ile Arg Asn Gly Thr Asp Pro His Lys Ala Asp Asn						
25		275	280			285	
	Leu Ser Ser Glu Met Ala Glu Tyr Ile Phe Ile Lys Pro Gly Thr Thr						
		290	295			300	
30	Asp Pro Leu Leu Cys Gly Asn Gly Ser Asp Ala Gly His Cys Pro Gly						
		305	310			315	320
	Gly Tyr Val Cys Leu Lys Thr Pro Asp Asn Pro Asp Phe Asn Tyr Thr						
		325	330			335	
35							
	Ser Phe Asp Ser Phe Ala Trp Ala Phe Leu Ser Leu Phe Arg Leu Met						
		340	345			350	
40	Thr Gln Asp Ser Trp Glu Arg Leu Tyr Gln Gln Thr Leu Arg Ala Ser						
		355	360			365	
	Gly Lys Met Tyr Met Val Phe Phe Val Leu Val Ile Phe Leu Gly Ser						
		370	375			380	
45	Phe Tyr Leu Val Asn Leu Ile Leu Ala Val Val Thr Met Ala Tyr Glu						
		385	390			395	400
	Glu Gln Ser Gln Ala Thr Ile Ala Glu Ile Glu Ala Lys Glu Lys Lys						
50		405	410			415	
	Phe Gln Glu Ala Leu Glu Val Leu Gln Lys Glu Gln Glu Val Leu Ala						
		420	425			430	
55	Ala Leu Gly Ile Asp Thr Thr Ser Leu Gln Ser His Ser Gly Ser Pro						
		435	440			445	
	Leu Ala Ser Lys Asn Ala Asn Glu Arg Arg Pro Arg Val Lys Ser Arg						
		450	455			460	
60	Val Ser Glu Gly Ser Thr Asp Asp Asn Arg Ser Pro Gln Ser Asp Pro						
		465	470			475	480

-80-

	Tyr	Asn	Gln	Arg	Arg	Met	Ser	Phe	Leu	Gly	Leu	Ser	Ser	Gly	Arg	Arg	
					485					490					495		
5	Arg	Ala	Ser	His	Gly	Ser	Val	Phe	His	Phe	Arg	Ala	Pro	Ser	Gln	Asp	
				500					505					510			
	Ile	Ser	Phe	Pro	Asp	Gly	Ile	Thr	Pro	Asp	Asp	Gly	Val	Phe	His	Gly	
			515					520					525				
10	Asp	Gln	Glu	Ser	Arg	Arg	Gly	Ser	Ile	Leu	Leu	Gly	Arg	Gly	Ala	Gly	
		530					535					540					
	Gln	Thr	Gly	Pro	Leu	Pro	Arg	Ser	Pro	Leu	Pro	Gln	Ser	Pro	Asn	Pro	
15		545				550					555				560		
	Gly	Arg	Arg	His	Gly	Glu	Glu	Gly	Gln	Leu	Gly	Val	Pro	Thr	Gly	Glu	
				565					570						575		
20	Leu	Thr	Ala	Gly	Ala	Pro	Glu	Gly	Pro	Ala	Leu	Asp	Thr	Thr	Gly	Gln	
				580					585					590			
	Lys	Ser	Phe	Leu	Ser	Ala	Gly	Tyr	Leu	Asn	Glu	Pro	Phe	Arg	Ala	Gln	
			595					600					605				
25	Arg	Ala	Met	Ser	Val	Val	Ser	Ile	Met	Thr	Ser	Val	Ile	Glu	Glu	Leu	
		610					615					620					
	Glu	Glu	Ser	Lys	Leu	Lys	Cys	Pro	Pro	Cys	Leu	Ile	Ser	Phe	Ala	Gln	
30		625				630					635					640	
	Lys	Tyr	Leu	Ile	Trp	Glu	Cys	Cys	Pro	Lys	Trp	Arg	Lys	Phe	Lys	Met	
				645					650						655		
35	Ala	Leu	Phe	Glu	Leu	Val	Thr	Asp	Pro	Phe	Ala	Glu	Leu	Thr	Ile	Thr	
				660				665						670			
	Leu	Cys	Ile	Val	Val	Asn	Thr	Val	Phe	Met	Ala	Met	Glu	His	Tyr	Pro	
			675				680						685				
40	Met	Thr	Asp	Ala	Phe	Asp	Ala	Met	Leu	Gln	Ala	Gly	Asn	Ile	Val	Phe	
		690				695						700					
	Thr	Val	Phe	Phe	Thr	Met	Glu	Met	Ala	Phe	Lys	Ile	Ile	Ala	Phe	Asp	
45		705			710						715					720	
	Pro	Tyr	Tyr	Tyr	Phe	Gln	Lys	Lys	Trp	Asn	Ile	Phe	Asp	Cys	Val	Ile	
				725					730					735			
50	Val	Thr	Val	Ser	Leu	Leu	Glu	Leu	Ser	Ala	Ser	Lys	Lys	Gly	Ser	Leu	
				740				745						750			
	Ser	Val	Leu	Arg	Ser	Leu	Arg	Leu	Ala	Leu	Asp	Thr	Thr	Gly	Gln	Lys	
			755				760						765				
55	Ser	Phe	Leu	Ser	Ala	Gly	Tyr	Leu	Asn	Glu	Pro	Phe	Arg	Ala	Gln	Arg	
		770				775						780					
	Ala	Met	Ser	Val	Val	Ser	He	Met	Thr	Ser	Val	Ile	Glu	Glu	Leu	Glu	
60		785			790						795				800		
	Glu	Ser	Lys	Leu	Lys	Cys	Pro	Pro	Cys	Leu	Ile	Ser	Phe	Ala	Gln	Lys	
				805						810					815		

-81-

	Tyr	Leu	Ile	Trp	Glu	Cys	Cys	Pro	Lys	Trp	Arg	Lys	Phe	Lys	Met	Ala	
				820					825					830			
5	Leu	Phe	Glu	Leu	Val	Thr	Asp	Pro	Phe	Ala	Glu	Leu	Thr	Ile	Thr	Leu	
			835					840					845				
	Cys	Ile	Val	Val	Asn	Thr	Val	Phe	Met	Ala	Met	Glu	His	Tyr	Pro	Met	
		850					855					860					
10	Thr	Asp	Ala	Phe	Asp	Ala	Met	Leu	Gln	Ala	Gly	Asn	Ile	Val	Phe	Thr	
	865					870					875					880	
	Val	Phe	Phe	Thr	Met	Glu	Met	Ala	Phe	Lys	Ile	Ile	Ala	Phe	Asp	Pro	
					885					890					895		
15	Tyr	Tyr	Tyr	Phe	Gln	Lys	Lys	Trp	Asn	Ile	Phe	Asp	Cys	Val	Ile	Val	
				900					905					910			
20	Thr	Val	Ser	Leu	Leu	Glu	Leu	Ser	Ala	Ser	Lys	Lys	Gly	Ser	Leu	Ser	
			915					920					925				
	Val	Leu	Arg	Ser	Leu	Arg	Leu	Leu	Arg	Val	Phe	Lys	Leu	Ala	Lys	Ser	
		930					935					940					
25	Trp	Pro	Thr	Leu	Asn	Thr	Leu	Ile	Lys	Ile	Ile	Gly	Asn	Ser	Val	Gly	
	945					950					955					960	
30	Ala	Leu	Gly	Asn	Leu	Thr	Phe	Ile	Leu	Ala	Ile	Ile	Val	Phe	Ile	Phe	
				965						970					975		
	Ala	Leu	Val	Gly	Lys	Gln	Leu	Leu	Ser	Glu	Asp	Tyr	Gly	Cys	Arg	Lys	
			980						985					990			
35	Asp	Gly	Val	Ser	Val	Trp	Asn	Gly	Glu	Lys	Leu	Arg	Trp	His	Met	Cys	
		995						1000					1005				
	Asp	Phe	Phe	His	Ser	Phe	Leu	Val	Val	Phe	Arg	Ile	Leu	Cys	Gly	Glu	
	1010					1015						1020					
40	Trp	Ile	Glu	Asn	Met	Trp	Val	Cys	Met	Glu	Val	Ser	Gln	Lys	Ser	Ile	
	1025					1030					1035					1040	
45	Cys	Leu	Ile	Leu	Phe	Leu	Thr	Val	Met	Val	Leu	Gly	Asn	Leu	Val	Val	
				1045						1050				1055			
	Leu	Asn	Leu	Phe	Ile	Ala	Leu	Leu	Leu	Asn	Ser	Phe	Ser	Ala	Asp	Asn	
			1060						1065					1070			
50	Leu	Thr	Ala	Pro	Glu	Asp	Asp	Gly	Glu	Val	Asn	Asn	Leu	Gln	Leu	Ala	
		1075						1080					1085				
	Leu	Ala	Arg	Ile	Gln	Val	Leu	Gly	His	Arg	Ala	Ser	Arg	Ala	Ile	Ala	
		1090				1095						1100					
55	Ser	Tyr	Ile	Ser	Ser	His	Cys	Arg	Phe	Arg	Trp	Pro	Lys	Val	Glu	Thr	
	1105					1110					1115				1120		
60	Gln	Leu	Gly	Met	Lys	Pro	Pro	Leu	Thr	Ser	Ser	Glu	Ala	Lys	Asn	His	
				1125						1130					1135		
	Ile	Ala	Thr	Asp	Ala	Val	Ser	Ala	Ala	Val	Gly	Asn	Leu	Thr	Lys	Pro	
			1140					1145					1150				

-82-

Ala Leu Ser Ser Pro Lys Glu Asn His Gly Asp Phe Ile Thr Asp Pro
 1155 1160 1165
 5 Asn Val Trp Val Ser Val Pro Ile Ala Glu Gly Glu Ser Asp Leu Asp
 1170 1175 1180
 Glu Leu Glu Glu Asp Met Glu Gln Ala Ser Gln Ser Ser Trp Gln Glu
 1185 1190 1195 1200
 10 Glu Asp Pro Lys Gly Gln Gln Glu Gln Leu Pro Gln Val Gln Lys Cys
 1205 1210 1215
 Glu Asn His Gln Ala Ala Arg Ser Pro Ala Ser Met Met Ser Ser Glu
 1220 1225 1230
 15 Asp Leu Ala Pro Tyr Leu Gly Glu Ser Trp Lys Arg Lys Asp Ser Pro
 1235 1240 1245
 Gln Val Pro Ala Glu Gly Val Asp Asp Thr Ser Ser Ser Glu Gly Ser
 1250 1255 1260
 20 Thr Val Asp Cys Pro Asp Pro Glu Glu Ile Leu Arg Lys Ile Pro Glu
 1265 1270 1275 1280
 Leu Ala Asp Asp Leu Asp Glu Pro Asp Asp Cys Phe Thr Glu Gly Cys
 1285 1290 1295
 25 Thr Arg Arg Cys Pro Cys Cys Asn Val Asn Thr Ser Lys Ser Pro Trp
 1300 1305 1310
 30 Ala Thr Gly Trp Gln Val Arg Lys Thr Cys Tyr Arg Ile Val Glu His
 1315 1320 1325
 Ser Trp Phe Glu Ser Phe Ile Ile Phe Met Ile Leu Leu Ser Ser Gly
 1330 1335 1340
 35 Ala Leu Ala Phe Glu Asp Asn Tyr Leu Glu Glu Lys Pro Arg Val Lys
 1345 1350 1355 1360
 Ser Val Leu Glu Tyr Thr Asp Arg Val Phe Thr Phe Ile Phe Val Phe
 1365 1370 1375
 40 Glu Met Leu Leu Lys Trp Val Ala Tyr Gly Phe Lys Lys Tyr Phe Thr
 1380 1385 1390
 45 Asn Ala Trp Cys Trp Leu Asp Phe Leu Ile Val Asn Ile Ser Leu Thr
 1395 1400 1405
 Ser Leu Ile Ala Lys Ile Leu Glu Tyr Ser Asp Val Ala Ser Ile Lys
 1410 1415 1420
 50 Ala Leu Arg Thr Leu Arg Ala Leu Arg Pro Leu Arg Ala Leu Ser Arg
 1425 1430 1435 1440
 Phe Glu Gly Met Arg Val Val Val Asp Ala Leu Val Gly Ala Ile Pro
 1445 1450 1455
 55 Ser Ile Met Asn Val Leu Leu Val Cys Leu Ile Phe Trp Leu Ile Phe
 1460 1465 1470
 60 Ser Ile Met Gly Val Asn Leu Phe Ala Gly Lys Phe Ser Lys Cys Val
 1475 1480 1485
 Asp Thr Arg Asn Asn Pro Phe Ser Asn Val Asn Ser Thr Met Val Asn
 1490 1495 1500

Asn Lys Ser Glu Cys His Asn Gln Asn Ser Thr Gly His Phe Phe Trp
 1505 1510 1515 1520
 5 Val Asn Val Lys Val Asn Phe Asp Asn Val Ala Met Gly Tyr Leu Ala
 1525 1530 1535
 Leu Leu Gln Val Ala Thr Phe Lys Gly Trp Met Asp Ile Met Tyr Ala
 1540 1545 1550
 10 Ala Val Asp Ser Gly Glu Ile Asn Ser Gln Pro Asn Trp Glu Asn Asn
 1555 1560 1565
 Leu Tyr Met Tyr Leu Tyr Phe Val Val Phe Ile Ile Phe Gly Gly Phe
 1570 1575 1580
 Phe Thr Leu Asn Leu Phe Val Gly Val Ile Ile Asp Asn Phe Asn Gln
 1585 1590 1595 1600
 20 Gln Lys Lys Lys Leu Gly Gly Gln Asp Ile Phe Met Thr Glu Glu Gln
 1605 1610 1615
 Lys Lys Tyr Tyr Asn Ala Met Lys Lys Leu Gly Ser Lys Lys Pro Gln
 1620 1625 1630
 25 Lys Pro Ile Pro Arg Pro Leu Asn Lys Tyr Gln Gly Phe Val Phe Asp
 1635 1640 1645
 Ile Val Thr Arg Gln Ala Phe Asp Ile Ile Ile Met Val Leu Ile Cys
 1650 1655 1660
 Leu Asn Met Ile Thr Met Met Val Glu Thr Asp Glu Gln Gly Glu Glu
 1665 1670 1675 1680
 35 Lys Thr Lys Val Leu Gly Arg Ile Asn Gln Phe Phe Val Ala Val Phe
 1685 1690 1695
 Thr Gly Glu Cys Val Met Lys Met Phe Ala Leu Arg Gln Tyr Tyr Phe
 1700 1705 1710
 40 Thr Asn Gly Trp Asn Val Phe Asp Phe Ile Val Val Ile Leu Ser Ile
 1715 1720 1725
 Gly Ser Leu Leu Phe Ser Ala Ile Leu Lys Ser Leu Glu Asn Tyr Phe
 1730 1735 1740
 Ser Pro Thr Leu Phe Arg Val Ile Arg Leu Ala Arg Ile Gly Arg Ile
 1745 1750 1755 1760
 50 Leu Arg Leu Ile Arg Ala Ala Lys Gly Ile Arg Thr Leu Leu Phe Ala
 1765 1770 1775
 Leu Met Met Ser Leu Pro Ala Leu Phe Asn Ile Gly Leu Leu Leu Phe
 1780 1785 1790
 55 Leu Val Met Phe Ile Tyr Ser Ile Phe Gly Met Ala Ser Phe Ala Asn
 1795 1800 1805
 Val Val Asp Glu Ala Gly Ile Asp Asp Met Phe Asn Phe Lys Thr Phe
 1810 1815 1820
 60 Gly Asn Ser Met Leu Cys Leu Phe Gln Ile Thr Thr Ser Ala Gly Trp
 1825 1830 1835 1840

-84-

Asp Gly Leu Leu Ser Pro Ile Leu Asn Thr Gly Pro Pro Tyr Cys Asp
 1845 1850 1855
 5 Pro Asn Leu Pro Asn Ser Asn Gly Ser Arg Gly Asn Cys Gly Ser Pro
 1860 1865 1870
 Ala Val Gly Ile Ile Phe Phe Thr Thr Tyr Ile Ile Ile Ser Phe Leu
 1875 1880 1885
 10 Ile Val Val Asn Met Tyr Ile Ala Val Ile Leu Glu Asn Phe Asn Val
 1890 1895 1900
 Ala Thr Glu Glu Ser Thr Glu Pro Leu Ser Glu Asp Asp Phe Asp Met
 1905 1910 1915 1920
 15 Phe Tyr Glu Thr Trp Glu Lys Phe Asp Pro Glu Ala Thr Gln Phe Ile
 1925 1930 1935
 Ala Phe Ser Ala Leu Ser Asp Phe Ala Asp Thr Leu Ser Gly Pro Leu
 1940 1945 1950
 20 Arg Ile Pro Lys Pro Asn Gln Asn Ile Leu Ile Gln Met Asp Leu Pro
 1955 1960 1965
 25 Leu Val Pro Gly Asp Lys Ile His Cys Leu Asp Ile Leu Phe Ala Phe
 1970 1975 1980
 Thr Lys Asn Val Leu Gly Glu Ser Gly Glu Leu Asp Ser Leu Lys Thr
 1985 1990 1995 2000
 Asn Met Glu Glu Lys Phe Met Ala Thr Asn Leu Ser Lys Ala Ser Tyr
 2005 2010 2015
 35 Glu Pro Ile Ala Thr Thr Leu Arg Trp Lys Gln Glu Asp Leu Ser Ala
 2020 2025 2030
 Thr Val Ile Gln Lys Ala Tyr Arg Ser Tyr Met Leu His Arg Ser Leu
 2035 2040 2045
 40 Thr Leu Ser Asn Thr Leu His Val Pro Arg Ala Glu Glu Asp Gly Val
 2050 2055 2060
 Ser Leu Pro Gly Glu Gly Tyr Ser Thr Phe Met Ala Asn Ser Gly Leu
 2065 2070 2075 2080
 45 Pro Asp Lys Ser Glu Thr Ala Ser Ala Thr Ser Phe Pro Pro Ser Tyr
 2085 2090 2095
 50 Asp Ser Val Thr Arg Gly Leu Ser Asp Arg Ala Asn Ile Asn Pro Ser
 2100 2105 2110
 Ser Ser Met Gln Asn Glu Asp Glu Val Ala Ala Lys Glu Gly Asn Ser
 2115 2120 2125
 55 Pro Gly Pro Gln
 2130

60 (2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6527 base pairs

(B) TYPE: nucleic acid

-85-

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 204..6077

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

5	TAGCTTGCTT CTGCTAATGC TACCCCAGGC CTTTAGACAG AGAACAGATG GCAGATGGAG	60
15	TTTCTTATTG CCATGCGCAA ACGCTGAGCC CACCTCATGA TCCCGGACCC CATGGTTTTC	120
	AGTAGACAAC CTGGGCTAAG AAGAGATCTC CGACCTTATA GAGCAGCAAA GAGTGTAAT	180
20	TCTTCCCCAA GAAGAATGAG AAG ATG GAG CTC CCC TTT GCG TCC GTG GGA	230
	Met Glu Leu Pro Phe Ala Ser Val Gly	
	1 5	
25	ACT ACC AAT TTC AGA CGG TTC ACT CCA GAG TCA CTG GCA GAG ATC GAG	278
	Thr Thr Asn Phe Arg Arg Phe Thr Pro Glu Ser Leu Ala Glu Ile Glu	
	10 15 20 25	
30	AAG CAG ATT GCT GCT CAC CGG GCA GCC AAG AAG GCC AGA ACC AAG CAC	326
	Lys Gln Ile Ala Ala His Arg Ala Ala Lys Lys Ala Arg Thr Lys His	
	30 35 40	
35	AGA GGA CAG GAG GAC AAG GGC GAG AAG CCC AGG CCT CAG CTG GAC TTG	374
	Arg Gly Gln Glu Asp Lys Gly Glu Lys Pro Arg Pro Gln Leu Asp Leu	
	45 50 55	
40	AAA GAC TGT AAC CAG CTG CCC AAG TTC TAT GGT GAG CTC CCA GCA GAA	422
	Lys Asp Cys Asn Gln Leu Pro Lys Phe Tyr Gly Glu Leu Pro Ala Glu	
	60 65 70	
45	CTG GTC GGG GAG CCC CTG GAG GAC CTA GAC CCT TTC TAC AGC ACA CAC	470
	Leu Val Gly Glu Pro Leu Glu Asp Leu Asp Pro Phe Tyr Ser Thr His	
	75 80 85	
50	CGG ACA TTC ATG GTG TTG AAT AAA AGC AGG ACC ATT TCC AGA TTC AGT	518
	Arg Thr Phe Met Val Leu Asn Lys Ser Arg Thr Ile Ser Arg Phe Ser	
	90 95 100 105	
55	GCC ACT TGG GCC CTG TGG CTC TTC AGT CCC TTC AAC CTG ATC AGA AGA	566
	Ala Thr Trp Ala Leu Trp Leu Phe Ser Pro Phe Asn Leu Ile Arg Arg	
	110 115 120	
60	ACA GCC ATC AAA GTG TCT GTC CAT TCC TGG TTC TCC ATA TTC ATC ACC	614
	Thr Ala Ile Lys Val Ser Val His Ser Trp Phe Ser Ile Phe Ile Thr	
	125 130 135	
65	ATC ACT ATT TTG GTC AAC TGC GTG TGC ATG ACC CGA ACT GAT CTT CCA	662
	Ile Thr Ile Leu Val Asn Cys Val Cys Met Thr Arg Thr Asp Leu Pro	
	140 145 150	
70	GAG AAA GTC GAG TAC GTC TTC ACT GTC ATT TAC ACC TTC GAG GCT CTG	710
	Glu Lys Val Glu Tyr Val Phe Thr Val Ile Tyr Thr Phe Glu Ala Leu	
	155 160 165	

-86-

	ATT AAG ATA CTG GCA AGA GGG TTT TGT CTA AAT GAG TTC ACT TAT CTT	758
	Ile Lys Ile Leu Ala Arg Gly Phe Cys Leu Asn Glu Phe Thr Tyr Leu	
	170 175 180 185	
5	CGA GAT CCG TGG AAC TGG CTG GAC TTC AGT GTC ATT ACC TTG GCG TAT	806
	Arg Asp Pro Trp Asn Trp Leu Asp Phe Ser Val Ile Thr Leu Ala Tyr	
	190 195 200	
10	GTG GGT GCA GCG ATA GAC CTC CGA GGA ATC TCA GGC CTG CGG ACA TTC	854
	Val Gly Ala Ala Ile Asp Leu Arg Gly Ile Ser Gly Leu Arg Thr Phe	
	205 210 215	
15	CGA GTT CTC AGA GCC CTG AAA ACT GTT TCT GTG ATC CCA GGA CTG AAG	902
	Arg Val Leu Arg Ala Leu Lys Thr Val Ser Val Ile Pro Gly Leu Lys	
	220 225 230	
20	GTC ATC GTG GGA GCC CTG ATC CAC TCA GTG AGG AAG CTG GCC GAC GTG	950
	Val Ile Val Gly Ala Leu Ile His Ser Val Arg Lys Leu Ala Asp Val	
	235 240 245	
25	ACT ATC CTC ACA GTC TTC TGC CTG AGC GTC TTC GCC TTG GTG GGC CTG	998
	Thr Ile Leu Thr Val Phe Cys Leu Ser Val Phe Ala Leu Val Gly Leu	
	250 255 260 265	
30	CAG CTC TTT AAG GGG AAC CTT AAG AAC AAA TGC ATC AGG AAC GGA ACA	1046
	Gln Leu Phe Lys Gly Asn Leu Lys Asn Lys Cys Ile Arg Asn Gly Thr	
	270 275 280	
35	GAT CCC CAC AAG GCT GAC AAC CTC TCA TCT GAA ATG GCA GAA TAC ATC	1094
	Asp Pro His Lys Ala Asp Asn Leu Ser Ser Glu Met Ala Glu Tyr Ile	
	285 290 295	
40	TTC ATC AAG CCT GGT ACT ACG GAT CCC TTA CTG TGC GGC AAT GGG TCT	1142
	Phe Ile Lys Pro Gly Thr Thr Asp Pro Leu Leu Cys Gly Asn Gly Ser	
	300 305 310	
45	GAT GCT GGT CAC TGC CCT GGA GGC TAT GTC TGC CTG AAA ACT CCT GAC	1190
	Asp Ala Gly His Cys Pro Gly Gly Tyr Val Cys Leu Lys Thr Pro Asp	
	315 320 325	
50	AAC CCG GAT TTT AAC TAC ACC AGC TTT GAT TCC TTT GCG TGG GCA TTC	1238
	Asn Pro Asp Phe Asn Tyr Thr Ser Phe Asp Ser Phe Ala Trp Ala Phe	
	330 335 340 345	
55	CTC TCA CTG TTC CGC CTC ATG ACG CAG GAC TCC TGG GAG CGC CTG TAC	1286
	Leu Ser Leu Phe Arg Leu Met Thr Gln Asp Ser Trp Glu Arg Leu Tyr	
	350 355 360	
60	CAG CAG ACA CTC CGG GCT TCT GGG AAA ATG TAC ATG GTC TTT TTC GTG	1334
	Gln Gln Thr Leu Arg Ala Ser Gly Lys Met Tyr Met Val Phe Phe Val	
	365 370 375	
65	CTG GTT ATT TTC CTT GGA TCG TTC TAC CTG GTC AAT TTG ATC TTG GCC	1382
	Leu Val Phe Leu Gly Ser Phe Tyr Leu Val Asn Leu Ile Leu Ala	
	380 385 390	
70	GTG GTC ACC ATG GCG TAT GAA GAG CAG AGC CAG GCA ACA ATT GCA GAA	1430
	Val Val Thr Met Ala Tyr Glu Glu Gln Ser Gln Ala Thr Ile Ala Glu	
	395 400 405	
75	ATC GAA GCC AAG GAA AAA AAG TTC CAG GAA GCC CTT GAG GTG CTG CAG	1478
	Ile Glu Ala Lys Glu Lys Lys Phe Gln Glu Ala Leu Glu Val Leu Gln	
	410 415 420 425	

	AAG GAA CAG GAG GTG CTG GCA GCC CTG GGG ATT GAC ACG ACC TCG CTC	1526
	Lys Glu Gln Glu Val Leu Ala Ala Leu Gly Ile Asp Thr Thr Ser Leu	
	430 435 440	
5	CAG TCC CAC AGT GGA TCA CCC TTA GCC TCC AAA AAC GCC AAT GAG AGA	1574
	Gln Ser His Ser Gly Ser Pro Leu Ala Ser Lys Asn Ala Asn Glu Arg	
	445 450 455	
10	AGA CCC AGG GTG AAA TCA AGG GTG TCA GAG GGC TCC ACG GAT GAC AAC	1622
	Arg Pro Arg Val Lys Ser Arg Val Ser Glu Gly Ser Thr Asp Asp Asn	
	460 465 470	
15	AGG TCA CCC CAA TCT GAC CCT TAC AAC CAG CGC AGG ATG TCT TTC CTA	1670
	Arg Ser Pro Gln Ser Asp Pro Tyr Asn Gln Arg Arg Met Ser Phe Leu	
	475 480 485	
20	GGC CTG TCT TCA GGA AGA CGC AGG GCT AGC CAC GGC AGT GTG TTC CAC	1718
	Gly Leu Ser Ser Gly Arg Arg Arg Ala Ser His Gly Ser Val Phe His	
	490 495 500 505	
25	TTC CGA GCG CCC AGC CAA GAC ATC TCA TTT CCT GAC GGG ATC ACC CCT	1766
	Phe Arg Ala Pro Ser Gln Asp Ile Ser Phe Pro Asp Gly Ile Thr Pro	
	510 515 520	
30	GAT GAT GGG GTC TTT CAC GGA GAC CAG GAA AGC CGT CGA GGT TCC ATA	1814
	Asp Asp Gly Val Phe His Gly Asp Gln Glu Ser Arg Arg Gly Ser Ile	
	525 530 535	
35	TTG CTG GGC AGG GGT GCT GGG CAG ACA GGT CCA CTC CCC AGG AGC CCA	1862
	Leu Leu Gly Arg Gly Ala Gly Gln Thr Gly Pro Leu Pro Arg Ser Pro	
	540 545 550	
40	CTG CCT CAG TCC CCC AAC CCT GGC CGT AGA CAT GGA GAA GAG GGA CAG	1910
	Leu Pro Gln Ser Pro Asn Pro Gly Arg Arg His Gly Glu Glu Gly Gln	
	555 560 565	
45	CTC GGA GTG CCC ACT GGT GAG CTT ACC GCT GGA GCG CCT GAA GGC CCG	1958
	Leu Gly Val Pro Thr Gly Glu Leu Thr Ala Gly Ala Pro Glu Gly Pro	
	570 575 580 585	
50	GCA CTC GAC ACT ACA GGG CAG AAG AGC TTC CTG TCT GCG GGC TAC TTG	2006
	Ala Leu Asp Thr Thr Gly Gln Lys Ser Phe Leu Ser Ala Gly Tyr Leu	
	590 595 600	
55	AAC GAA CCT TTC CGA GCA CAG AGG GCC ATG AGC GTT GTC AGT ATC ATG	2054
	Asn Glu Pro Phe Arg Ala Gln Arg Ala Met Ser Val Val Ser Ile Met	
	605 610 615	
60	ACT TCT GTC ATT GAG GAG CTT GAA GAG TCT AAG CTG AAG TGC CCA CCC	2102
	Thr Ser Val Ile Glu Glu Leu Glu Ser Lys Leu Lys Cys Pro Pro	
	620 625 630	
65	TGC TTG ATC AGC TTC GCT CAG AAG TAT CTG ATC TGG GAG TGC TGC CCC	2150
	Cys Leu Ile Ser Phe Ala Gln Lys Tyr Leu Ile Trp Glu Cys Cys Pro	
	635 640 645	
70	AAG TGG AGG AAG TTC AAG ATG GCG CTG TTC GAG CTG GTG ACT GAC CCC	2198
	Lys Trp Arg Lys Phe Lys Met Ala Leu Phe Glu Leu Val Thr Asp Pro	
	650 655 660 665	
75	TTC GCA GAG CTT ACC ATC ACC CTC TGC ATC GTG GTG AAC ACC GTC TTC	2246
	Phe Ala Glu Leu Thr Ile Thr Leu Cys Ile Val Val Asn Thr Val Phe	
	670 675 680	

-88-

	ATG	GCC	ATG	GAG	CAC	TAC	CCC	ATG	ACC	GAT	GCC	TTC	GAT	GCC	ATG	CTT	2294
	Met	Ala	Met	Glu	His	Tyr	Pro	Met	Thr	Asp	Ala	Phe	Asp	Ala	Met	Leu	
				685					690					695			
5	CAA	GCC	GGC	AAC	ATT	GTC	TTC	ACC	GTG	TTT	TTC	ACA	ATG	GAG	ATG	GCC	2342
	Gln	Ala	Gly	Asn	Ile	Val	Phe	Thr	Val	Phe	Phe	Thr	Met	Glu	Met	Ala	
			700					705					710				
10	TTC	AAG	ATC	ATT	GCC	TTC	GAC	CCC	TAC	TAT	TAC	TTC	CAG	AAG	AAG	TGG	2390
	Phe	Lys	Ile	Ile	Ala	Phe	Asp	Pro	Tyr	Tyr	Tyr	Phe	Gln	Lys	Lys	Trp	
		715					720					725					
15	AAT	ATC	TTC	GAC	TGT	GTC	ATC	GTC	ACC	GTG	AGC	CTT	CTG	GAG	CTG	AGT	2438
	Asn	Ile	Phe	Asp	Cys	Val	Ile	Val	Thr	Val	Ser	Leu	Leu	Glu	Leu	Ser	
	730					735					740					745	
20	GCA	TCC	AAG	AAG	GGC	AGC	CTG	TCT	GTG	CTC	CGT	TCC	TTA	CGC	TTG	CTG	2486
	Ala	Ser	Lys	Lys	Gly	Ser	Leu	Ser	Val	Leu	Arg	Ser	Leu	Arg	Leu	Leu	
					750					755					760		
25	CGG	GTC	TTC	AAG	CTG	GCC	AAG	TCC	TGG	CCC	ACC	CTG	AAC	ACC	CTC	ATC	2534
	Arg	Val	Phe	Lys	Leu	Ala	Lys	Ser	Trp	Pro	Thr	Leu	Asn	Thr	Leu	Ile	
				765					770					775			
30	AAG	ATC	ATC	GGG	AAC	TCA	GTG	GGG	GCC	CTG	GGC	AAC	CTG	ACC	TTT	ATC	2582
	Lys	Ile	Ile	Gly	Asn	Ser	Val	Gly	Ala	Leu	Gly	Asn	Leu	Thr	Phe	Ile	
				780				785					790				
35	CTG	GCC	ATC	ATC	GTC	TTC	ATC	TTC	GCC	CTG	GTC	GGA	AAG	CAG	CTT	CTC	2630
	Leu	Ala	Ile	Ile	Val	Phe	Ile	Phe	Ala	Leu	Val	Gly	Lys	Gln	Leu	Leu	
		795					800					805					
40	TCA	GAG	GAC	TAC	GGG	TGC	CGC	AAG	GAC	GGC	GTC	TCC	GTG	TGG	AAC	GGC	2678
	Ser	Glu	Asp	Tyr	Gly	Cys	Arg	Lys	Asp	Gly	Val	Ser	Val	Trp	Asn	Gly	
	810					815					820					825	
45	GAG	AAG	CTC	CGC	TGG	CAC	ATG	TGT	GAC	TTC	TTC	CAT	TCC	TTC	CTG	GTC	2726
	Glu	Lys	Leu	Arg	Trp	His	Met	Cys	Asp	Phe	Phe	His	Ser	Phe	Leu	Val	
					830					835					840		
50	GTC	TTC	CGA	ATC	CTC	TGC	GGG	GAG	TGG	ATC	GAG	AAC	ATG	TGG	GTC	TGC	2774
	Val	Phe	Arg	Ile	Leu	Cys	Gly	Glu	Trp	Ile	Glu	Asn	Met	Trp	Val	Cys	
				845					850					855			
55	ATG	GAG	GTC	AGC	CAG	AAA	TCC	ATC	TGC	CTC	ATC	CTC	TTC	TTG	ACT	GTG	2822
	Met	Glu	Val	Ser	Gln	Lys	Ser	Ile	Cys	Leu	Ile	Leu	Phe	Leu	Thr	Val	
			860					865					870				
60	ATG	GTG	CTG	GGC	AAC	CTA	GTG	GTG	CTC	AAC	CTT	TTC	ATC	GCT	TTA	CTG	2870
	Met	Val	Leu	Gly	Asn	Leu	Val	Val	Leu	Asn	Leu	Phe	Ile	Ala	Leu	Leu	
		875					880					885					
65	CTG	AAC	TCC	TTC	AGC	GCG	GAC	AAC	CTC	ACG	GCT	CCA	GAG	GAT	GAC	GGG	2918
	Leu	Asn	Ser	Phe	Ser	Ala	Asp	Asn	Leu	Thr	Ala	Pro	Glu	Asp	Asp	Gly	
	890					895					900					905	
70	GAG	GTG	AAC	AAC	TTG	CAG	TTA	GCA	CTG	GCC	AGG	ATC	CAG	GTA	CTT	GGC	2966
	Glu	Val	Asn	Asn	Leu	Gln	Leu	Ala	Leu	Ala	Arg	Ile	Gln	Val	Leu	Gly	
					910				915						920		
75	CAT	CGG	GCC	AGC	AGG	GCC	ATC	GCC	AGT	TAC	ATC	AGC	AGC	CAC	TGC	CGA	3014
	His	Arg	Ala	Ser	Arg	Ala	Ile	Ala	Ser	Tyr	Ile	Ser	Ser	His	Cys	Arg	
				925					930					935			

	TTC	CGC	TGG	CCC	AAG	GTG	GAG	ACC	CAG	CTG	GGC	ATG	AAG	CCC	CCA	CTC	3062
	Phe	Arg	Trp	Pro	Lys	Val	Glu	Thr	Gln	Leu	Gly	Met	Lys	Pro	Pro	Leu	
			940					945					950				
5	ACC	AGC	TCA	GAG	GCC	AAG	AAC	CAC	ATT	GCC	ACT	GAT	GCT	GTC	AGT	GCT	3110
	Thr	Ser	Ser	Glu	Ala	Lys	Asn	His	Ile	Ala	Thr	Asp	Ala	Val	Ser	Ala	
		955					960					965					
10	GCA	GTG	GGG	AAC	CTG	ACA	AAG	CCA	GCT	CTC	AGT	AGC	CCC	AAG	GAG	AAT	3158
	Ala	Val	Gly	Asn	Leu	Thr	Lys	Pro	Ala	Leu	Ser	Ser	Pro	Lys	Glu	Asn	
		970					975					980				985	
15	CAC	GGG	GAC	TTC	ATC	ACT	GAT	CCC	AAC	GTG	TGG	GTC	TCT	GTG	CCC	ATT	3206
	His	Gly	Asp	Phe	Ile	Thr	Asp	Pro	Asn	Val	Trp	Val	Ser	Val	Pro	Ile	
					990					995					1000		
20	GCT	GAG	GGG	GAA	TCT	GAC	CTC	GAC	GAG	CTC	GAG	GAA	GAT	ATG	GAG	CAG	3254
	Ala	Glu	Gly	Glu	Ser	Asp	Leu	Asp	Glu	Leu	Glu	Glu	Asp	Met	Glu	Gln	
				1005					1010					1015			
25	GCT	TCG	CAG	AGC	TCC	TGG	CAG	GAA	GAG	GAC	CCC	AAG	GGA	CAG	CAG	GAG	3302
	Ala	Ser	Gln	Ser	Ser	Trp	Gln	Glu	Glu	Asp	Pro	Lys	Gly	Gln	Gln	Glu	
			1020					1025					1030				
30	CAG	TTG	CCA	CAA	GTC	CAA	AAG	TGT	GAA	AAC	CAC	CAG	GCA	GCC	AGA	AGC	3350
	Gln	Leu	Pro	Gln	Val	Gln	Lys	Cys	Glu	Asn	His	Gln	Ala	Ala	Arg	Ser	
		1035					1040					1045					
35	CCA	GCC	TCC	ATG	ATG	TCC	TCT	GAG	GAC	CTG	GCT	CCA	TAC	CTG	GGT	GAG	3398
	Pro	Ala	Ser	Met	Met	Ser	Ser	Glu	Asp	Leu	Ala	Pro	Tyr	Leu	Gly	Glu	
		1050					1055				1060					1065	
40	AGC	TGG	AAG	AGG	AAG	GAT	AGC	CCT	CAG	GTC	CCT	GCC	GAG	GGA	GTG	GAT	3446
	Ser	Trp	Lys	Arg	Lys	Asp	Ser	Pro	Gln	Val	Pro	Ala	Glu	Gly	Val	Asp	
					1070					1075					1080		
45	GAC	ACG	AGC	TCC	TCT	GAG	GGC	AGC	ACG	GTG	GAC	TGC	CCG	GAC	CCA	GAG	3494
	Asp	Thr	Ser	Ser	Ser	Glu	Gly	Ser	Thr	Val	Asp	Cys	Pro	Asp	Pro	Glu	
				1085				1090						1095			
50	GAA	ATC	CTG	AGG	AAG	ATC	CCC	GAG	CTG	GCA	GAT	GAC	CTG	GAC	GAG	CCC	3542
	Glu	Ile	Leu	Arg	Lys	Ile	Pro	Glu	Leu	Ala	Asp	Asp	Leu	Asp	Glu	Pro	
			1100					1105					1110				
55	GAT	GAC	TGT	TTC	ACA	GAA	GGC	TGC	ACT	CGC	CGC	TGT	CCC	TGC	TGC	AAC	3590
	Asp	Asp	Cys	Phe	Thr	Glu	Gly	Cys	Thr	Arg	Arg	Cys	Pro	Cys	Cys	Asn	
		1115					1120					1125					
60	GTG	AAT	ACT	AGC	AAG	TCT	CCT	TGG	GCC	ACA	GGC	TGG	CAG	GTG	CGC	AAG	3638
	Val	Asn	Thr	Ser	Lys	Ser	Pro	Trp	Ala	Thr	Gly	Trp	Gln	Val	Arg	Lys	
		1130					1135					1140				1145	
65	ACC	TGC	TAC	CGC	ATC	GTG	GAG	CAC	AGC	TGG	TTT	GAG	AGT	TTC	ATC	ATC	3686
	Thr	Cys	Tyr	Arg	Ile	Val	Glu	His	Ser	Trp	Phe	Glu	Ser	Phe	Ile	Ile	
					1150					1155					1160		
70	TTC	ATG	ATC	CTG	CTC	AGC	AGT	GGA	GCG	CTG	GCC	TTT	GAG	GAT	AAC	TAC	3734
	Phe	Met	Ile	Leu	Leu	Ser	Ser	Gly	Ala	Leu	Ala	Phe	Glu	Asp	Asn	Tyr	
				1165				1170					1175				
75	CTG	GAA	GAG	AAA	CCC	CGA	GTG	AAG	TCC	GTG	CTG	GAG	TAC	ACT	GAC	CGA	3782
	Leu	Glu	Glu	Lys	Pro	Arg	Val	Lys	Ser	Val	Leu	Glu	Tyr	Thr	Asp	Arg	
			1180					1185					1190				

-90-

	GTG TTC ACC TTC ATC TTC GTC TTT GAG ATG CTG CTC AAG TGG GTA GCC	3830
	Val Phe Thr Phe Ile Phe Val Phe Glu Met Leu Leu Lys Trp Val Ala	
	1195 1200 1205	
5	TAT GGC TTC AAA AAG TAT TTC ACC AAT GCC TGG TGC TGG CTG GAC TTC	3878
	Tyr Gly Phe Lys Lys Tyr Phe Thr Asn Ala Trp Cys Trp Leu Asp Phe	
	1210 1215 1220 1225	
10	CTC ATT GTG AAC ATC TCC CTG ACA AGC CTC ATA GCG AAG ATC CTT GAG	3926
	Leu Ile Val Asn Ile Ser Leu Thr Ser Leu Ile Ala Lys Ile Leu Glu	
	1230 1235 1240	
15	TAT TCC GAC GTG GCG TCC ATC AAA GCC CTT CGG ACT CTC CGT GCC CTC	3974
	Tyr Ser Asp Val Ala Ser Ile Lys Ala Leu Arg Thr Leu Arg Ala Leu	
	1245 1250 1255	
20	CGA CCG CTG CGG GCT CTG TCT CGA TTC GAA GGC ATG AGG GTA GTG GTG	4022
	Arg Pro Leu Arg Ala Leu Ser Arg Phe Glu Gly Met Arg Val Val Val	
	1260 1265 1270	
25	GAT GCC CTC GTG GGC GCC ATC CCC TCC ATC ATG AAC GTC CTC CTC GTC	4070
	Asp Ala Leu Val Gly Ala Ile Pro Ser Ile Met Asn Val Leu Leu Val	
	1275 1280 1285	
30	TGC CTC ATC TTC TGG CTC ATC TTC AGC ATC ATG GGC GTG AAC CTC TTC	4118
	Cys Leu Ile Phe Trp Leu Ile Phe Ser Ile Met Gly Val Asn Leu Phe	
	1290 1295 1300 1305	
35	GCC GGG AAA TTT TCG AAG TGC GTC GAC ACC AGA AAT AAC CCA TTT TCC	4166
	Ala Gly Lys Phe Ser Lys Cys Val Asp Thr Arg Asn Asn Pro Phe Ser	
	1310 1315 1320	
40	AAC GTG AAT TCG ACG ATG GTG AAT AAC AAG TCC GAG TGT CAC AAT CAA	4214
	Asn Val Asn Ser Thr Met Val Asn Asn Lys Ser Glu Cys His Asn Gln	
	1325 1330 1335	
45	AAC AGC ACC GGC CAC TTC TTC TGG GTC AAC GTC AAA GTC AAC TTC GAC	4262
	Asn Ser Thr Gly His Phe Phe Trp Val Asn Val Lys Val Asn Phe Asp	
	1340 1345 1350	
50	AAC GTC GCT ATG GGC TAC CTC GCA CTT CTT CAG GTG GCA ACC TTC AAA	4310
	Asn Val Ala Met Gly Tyr Leu Ala Leu Leu Gln Val Ala Thr Phe Lys	
	1355 1360 1365	
55	GGC TGG ATG GAC ATA ATG TAT GCA GCT GTT GAT TCC GGA GAG ATC AAC	4358
	Gly Trp Met Asp Ile Met Tyr Ala Ala Val Asp Ser Gly Glu Ile Asn	
	1370 1375 1380 1385	
60	AGT CAG CCT AAC TGG GAG AAC AAC TTG TAC ATG TAC CTG TAC TTC GTC	4406
	Ser Gln Pro Asn Trp Glu Asn Asn Leu Tyr Met Tyr Leu Tyr Phe Val	
	1390 1395 1400	
65	GTT TTC ATC ATT TTC GGT GGC TTC TTC ACG CTG AAT CTC TTT GTT GGG	4454
	Val Phe Ile Ile Phe Gly Gly Phe Phe Thr Leu Asn Leu Phe Val Gly	
	1405 1410 1415	
70	GTC ATA ATC GAC AAC TTC AAC CAA CAG AAA AAA AAG CTA GGA GGC CAG	4502
	Val Ile Ile Asp Asn Phe Asn Gln Gln Lys Lys Lys Leu Gly Gly Gln	
	1420 1425 1430	
75	GAC ATC TTC ATG ACA GAA GAG CAG AAG AAG TAC TAC AAT GCC ATG AAG	4550
	Asp Ile Phe Met Thr Glu Glu Gln Lys Lys Tyr Tyr Asn Ala Met Lys	
	1435 1440 1445	

	AAG CTG GGC TCC AAG AAA CCC CAG AAG CCC ATC CCA CGG CCC CTG AAT	4598
	Lys Leu Gly Ser Lys Lys Pro Gln Lys Pro Ile Pro Arg Pro Leu Asn	
	1450 1455 1460 1465	
5	AAG TAC CAA GGC TTC GTG TTT GAC ATC GTG ACC AGG CAA GCC TTT GAC	4646
	Lys Tyr Gln Gly Phe Val Phe Asp Ile Val Thr Arg Gln Ala Phe Asp	
	1470 1475 1480	
10	ATC ATC ATC ATG GTT CTC ATC TGC CTC AAC ATG ATC ACC ATG ATG GTG	4694
	Ile Ile Ile Met Val Leu Ile Cys Leu Asn Met Ile Thr Met Met Val	
	1485 1490 1495	
15	GAG ACC GAC GAG CAG GGC GAG GAG AAG ACG AAG GTT CTG GGC AGA ATC	4742
	Glu Thr Asp Glu Gln Gly Glu Glu Lys Thr Lys Val Leu Gly Arg Ile	
	1500 1505 1510	
	AAC CAG TTC TTT GTG GCC GTC TTC ACG GGC GAG TGT GTG ATG AAG ATG	4790
	Asn Gln Phe Phe Val Ala Val Phe Thr Gly Glu Cys Val Met Lys Met	
	1515 1520 1525	
20	TTC GCC CTG CGA CAG TAC TAC TTC ACC AAC GGC TGG AAC GTG TTC GAC	4838
	Phe Ala Leu Arg Gln Tyr Tyr Phe Thr Asn Gly Trp Asn Val Phe Asp	
	1530 1535 1540 1545	
25	TTC ATA GTG GTG ATC CTG TCC ATT GGG AGT CTG CTG TTT TCT GCA ATC	4886
	Phe Ile Val Val Ile Leu Ser Ile Gly Ser Leu Leu Phe Ser Ala Ile	
	1550 1555 1560	
30	CTT AAG TCA CTG GAA AAC TAC TTC TCC CCG ACG CTC TTC CGG GTC ATC	4934
	Leu Lys Ser Leu Glu Asn Tyr Phe Ser Pro Thr Leu Phe Arg Val Ile	
	1565 1570 1575	
35	CGT CTG GCC AGG ATC GGC CGC ATC CTC AGG CTG ATC CGA GCA GCC AAG	4982
	Arg Leu Ala Arg Ile Gly Arg Ile Leu Arg Leu Ile Arg Ala Ala Lys	
	1580 1585 1590	
40	GGG ATT CGC ACG CTG CTC TTC GCC CTC ATG ATG TCC CTG CCC GCC CTC	5030
	Gly Ile Arg Thr Leu Leu Phe Ala Leu Met Met Ser Leu Pro Ala Leu	
	1595 1600 1605	
45	TTC AAC ATC GGC CTC CTC CTC TTC CTC GTC ATG TTC ATC TAC TCC ATC	5078
	Phe Asn Ile Gly Leu Leu Leu Phe Leu Val Met Phe Ile Tyr Ser Ile	
	1610 1615 1620 1625	
50	TTC GGC ATG GCC AGC TTC GCT AAC GTC GTG GAC GAG GCC GGC ATC GAC	5126
	Phe Gly Met Ala Ser Phe Ala Asn Val Val Asp Glu Ala Gly Ile Asp	
	1630 1635 1640	
55	GAC ATG TTC AAC TTC AAG ACC TTT GGC AAC AGC ATG CTG TGC CTG TTC	5174
	Asp Met Phe Asn Phe Lys Thr Phe Gly Asn Ser Met Leu Cys Leu Phe	
	1645 1650 1655	
60	CAG ATC ACC ACC TCG GCC GGC TGG GAC GGC CTC CTC AGC CCC ATC CTC	5222
	Gln Ile Thr Thr Ser Ala Gly Trp Asp Gly Leu Leu Ser Pro Ile Leu	
	1660 1665 1670	
	AAC ACG GGC CCT CCC TAC TGC GAC CCC AAC CTG CCC AAC AGC AAC GGC	5270
	Asn Thr Gly Pro Pro Tyr Cys Asp Pro Asn Leu Pro Asn Ser Asn Gly	
	1675 1680 1685	
60	TCC CGG GGC AAC TGC GGC AGC CCG GCG GTG GGC ATC ATC TTC TTC ACC	5318
	Ser Arg Gly Asn Cys Gly Ser Pro Ala Val Gly Ile Ile Phe Phe Thr	
	1690 1695 1700 1705	

-92-

	ACC TAC ATC ATC ATC TCC TTC CTC ATC GTG GTC AAC ATG TAC ATC GCA	5366
	Thr Tyr Ile Ile Ile Ser Phe Leu Ile Val Val Asn Met Tyr Ile Ala	
	1710 1715 1720	
5	GTG ATT CTG GAG AAC TTC AAC GTA GCC ACC GAG GAG AGC ACG GAG CCC	5414
	Val Ile Leu Glu Asn Phe Asn Val Ala Thr Glu Glu Ser Thr Glu Pro	
	1725 1730 1735	
10	CTG AGC GAG GAC GAC TTC GAC ATG TTC TAT GAG ACC TGG GAG AAG TTC	5462
	Leu Ser Glu Asp Asp Phe Asp Met Phe Tyr Glu Thr Trp Glu Lys Phe	
	1740 1745 1750	
15	GAC CCG GAG GCC ACC CAG TTC ATT GCC TTT TCT GCC CTC TCA GAC TTC	5510
	Asp Pro Glu Ala Thr Gln Phe Ile Ala Phe Ser Ala Leu Ser Asp Phe	
	1755 1760 1765	
20	GCG GAC ACG CTC TCC GGC CCT CTT AGA ATC CCC AAA CCC AAC CAG AAT	5558
	Ala Asp Thr Leu Ser Gly Pro Leu Arg Ile Pro Lys Pro Asn Gln Asn	
	1770 1775 1780 1785	
25	ATA TTA ATC CAG ATG GAC CTG CCG TTG GTC CCC GGG GAT AAG ATC CAC	5606
	Ile Leu Ile Gln Met Asp Leu Pro Leu Val Pro Gly Asp Lys Ile His	
	1790 1795 1800	
30	TGT CTG GAC ATC CTT TTT GCC TTC ACA AAG AAC GTC TTG GGA GAA TCC	5654
	Cys Leu Asp Ile Leu Phe Ala Phe Thr Lys Asn Val Leu Gly Glu Ser	
	1805 1810 1815	
35	GGG GAG TTG GAC TCC CTG AAG ACC AAT ATG GAA GAG AAG TTT ATG GCG	5702
	Gly Glu Leu Asp Ser Leu Lys Thr Asn Met Glu Glu Lys Phe Met Ala	
	1820 1825 1830	
40	ACC AAT CTC TCC AAA GCA TCC TAT GAA CCA ATA GCC ACC ACC CTC CGG	5750
	Thr Asn Leu Ser Lys Ala Ser Tyr Glu Pro Ile Ala Thr Thr Leu Arg	
	1835 1840 1845	
45	TGG AAG CAG GAA GAC CTC TCA GCC ACA GTC ATT CAA AAG GCC TAC CGG	5798
	Trp Lys Gln Glu Asp Leu Ser Ala Thr Val Ile Gln Lys Ala Tyr Arg	
	1850 1855 1860 1865	
50	AGC TAC ATG CTG CAC CGC TCC TTG ACA CTC TCC AAC ACC CTG CAT GTG	5846
	Ser Tyr Met Leu His Arg Ser Leu Thr Leu Ser Asn Thr Leu His Val	
	1870 1875 1880	
55	CCC AGG GCT GAG GAG GAT GGC GTG TCA CTT CCC GGG GAA GGC TAC AGT	5894
	Pro Arg Ala Glu Glu Asp Gly Val Ser Leu Pro Gly Glu Gly Tyr Ser	
	1885 1890 1895	
60	ACA TTC ATG GCA AAC AGT GGA CTC CCG GAC AAA TCA GAA ACT GCC TCT	5942
	Thr Phe Met Ala Asn Ser Gly Leu Pro Asp Lys Ser Glu Thr Ala Ser	
	1900 1905 1910	
65	GCT ACG TCT TTC CCG CCA TCC TAT GAC AGT GTC ACC AGG GGC CTG AGT	5990
	Ala Thr Ser Phe Pro Pro Ser Tyr Asp Ser Val Thr Arg Gly Leu Ser	
	1915 1920 1925	
70	GAC CGG GCC AAC ATT AAC CCA TCT AGC TCA ATG CAA AAT GAA GAT GAG	6038
	Asp Arg Ala Asn Ile Asn Pro Ser Ser Ser Met Gln Asn Glu Asp Glu	
	1930 1935 1940 1945	
75	GTC GCT GCT AAG GAA GGA AAC AGC CCT GGA CCT CAG TGAAGGCACT	6084
	Val Ala Ala Lys Glu Gly Asn Ser Pro Gly Pro Gln	
	1950 1955	

CAGGCATGCA CAGGGCAGGT TCCAATGTCT TTCTCTGCTG TACTAACTCC TTCCCTCTGG 6144
 AGGTGGCACC AACCTCCAGC CTCCACCAAT GCATGTCAC TGGTCATGGTG TCAGAACTGA 6204
 5 ATGGGGACAT CCTTGAGAAA GCCCCACCC CAATAGGAAT CAAAAGCCAA GGATACTCCT 6264
 CCATTCTGAC GTCCCTTCCG AGTTCCCAGA AGATGTCATT GCTCCCTTCT GTTTGTGACC 6324
 AGAGACGTGA TTCACCAACT TCTCGGAGCC AGAGACACAT AGCAAAGACT TTTCTGCTGG 6384
 10 TGTCGGGCAG TCTTAGAGAA GTCACGTAGG GGTGTTGACT GAGAATTAGG GTTTCATGA 6444
 CTGCATGCTC ACAGCTGCCG GACAATACCT GTGAGTCGGC CATTAAAATT AATATTTTTA 6504
 15 AAGTTAAAAA AAAAAAAAAA AAA 6527

(2) INFORMATION FOR SEQ ID NO:8:

20 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1957 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Glu Leu Pro Phe Ala Ser Val Gly Thr Thr Asn Phe Arg Arg Phe
 30 1 5 10 15
 Thr Pro Glu Ser Leu Ala Glu Ile Glu Lys Gln Ile Ala Ala His Arg
 20 25 30
 35 Ala Ala Lys Lys Ala Arg Thr Lys His Arg Gly Gln Glu Asp Lys Gly
 35 40 45
 Glu Lys Pro Arg Pro Gln Leu Asp Leu Lys Asp Cys Asn Gln Leu Pro
 50 55 60
 40 Lys Phe Tyr Gly Glu Leu Pro Ala Glu Leu Val Gly Glu Pro Leu Glu
 65 70 75 80
 Asp Leu Asp Pro Phe Tyr Ser Thr His Arg Thr Phe Met Val Leu Asn
 45 85 90 95
 Lys Ser Arg Thr Ile Ser Arg Phe Ser Ala Thr Trp Ala Leu Trp Leu
 100 105 110
 50 Phe Ser Pro Phe Asn Leu Ile Arg Arg Thr Ala Ile Lys Val Ser Val
 115 120 125
 His Ser Trp Phe Ser Ile Phe Ile Thr Ile Thr Ile Leu Val Asn Cys
 130 135 140
 55 Val Cys Met Thr Arg Thr Asp Leu Pro Glu Lys Val Glu Tyr Val Phe
 145 150 155 160
 Thr Val Ile Tyr Thr Phe Glu Ala Leu Ile Lys Ile Leu Ala Arg Gly
 60 165 175
 Phe Cys Leu Asn Glu Phe Thr Tyr Leu Arg Asp Pro Trp Asn Trp Leu
 180 185 190

	Asp	Phe	Ser	Val	Ile	Thr	Leu	Ala	Tyr	Val	Gly	Ala	Ala	Ile	Asp	Leu	
			195					200					205				
5	Arg	Gly	Ile	Ser	Gly	Leu	Arg	Thr	Phe	Arg	Val	Leu	Arg	Ala	Leu	Lys	
		210					215					220					
	Thr	Val	Ser	Val	Ile	Pro	Gly	Leu	Lys	Val	Ile	Val	Gly	Ala	Leu	Ile	
	225					230					235					240	
10	His	Ser	Val	Arg	Lys	Leu	Ala	Asp	Val	Thr	Ile	Leu	Thr	Val	Phe	Cys	
					245					250					255		
	Leu	Ser	Val	Phe	Ala	Leu	Val	Gly	Leu	Gln	Leu	Phe	Lys	Gly	Asn	Leu	
				260					265					270			
15	Lys	Asn	Lys	Cys	Ile	Arg	Asn	Gly	Thr	Asp	Pro	His	Lys	Ala	Asp	Asn	
			275					280					285				
	Leu	Ser	Ser	Glu	Met	Ala	Glu	Tyr	Ile	Phe	Ile	Lys	Pro	Gly	Thr	Thr	
20		290					295					300					
	Asp	Pro	Leu	Leu	Cys	Gly	Asn	Gly	Ser	Asp	Ala	Gly	His	Cys	Pro	Gly	
	305					310					315					320	
25	Gly	Tyr	Val	Cys	Leu	Lys	Thr	Pro	Asp	Asn	Pro	Asp	Phe	Asn	Tyr	Thr	
					325					330					335		
	Ser	Phe	Asp	Ser	Phe	Ala	Trp	Ala	Phe	Leu	Ser	Leu	Phe	Arg	Leu	Met	
				340					345					350			
30																	
	Thr	Gln	Asp	Ser	Trp	Glu	Arg	Leu	Tyr	Gln	Gln	Thr	Leu	Arg	Ala	Ser	
			355					360					365				
35	Gly	Lys	Met	Tyr	Met	Val	Phe	Phe	Val	Leu	Val	Ile	Phe	Leu	Gly	Ser	
		370					375					380					
	Phe	Tyr	Leu	Val	Asn	Leu	Ile	Leu	Ala	Val	Val	Thr	Met	Ala	Tyr	Glu	
	385					390					395					400	
40	Glu	Gln	Ser	Gln	Ala	Thr	Ile	Ala	Glu	Ile	Glu	Ala	Lys	Glu	Lys	Lys	
					405					410					415		
	Phe	Gln	Glu	Ala	Leu	Glu	Val	Leu	Gln	Lys	Glu	Gln	Glu	Val	Leu	Ala	
45				420					425					430			
	Ala	Leu	Gly	Ile	Asp	Thr	Thr	Ser	Leu	Gln	Ser	His	Ser	Gly	Ser	Pro	
			435					440					445				
50	Leu	Ala	Ser	Lys	Asn	Ala	Asn	Glu	Arg	Arg	Pro	Arg	Val	Lys	Ser	Arg	
		450					455					460					
	Val	Ser	Glu	Gly	Ser	Thr	Asp	Asp	Asn	Arg	Ser	Pro	Gln	Ser	Asp	Pro	
	465					470					475					480	
55	Tyr	Asn	Gln	Arg	Arg	Met	Ser	Phe	Leu	Gly	Leu	Ser	Ser	Gly	Arg	Arg	
					485					490					495		
	Arg	Ala	Ser	His	Gly	Ser	Val	Phe	His	Phe	Arg	Ala	Pro	Ser	Gln	Asp	
60				500					505					510			
	Ile	Ser	Phe	Pro	Asp	Gly	Ile	Thr	Pro	Asp	Asp	Gly	Val	Phe	His	Gly	
				515				520					525				

-95-

Asp Gln Glu Ser Arg Arg Gly Ser Ile Leu Leu Gly Arg Gly Ala Gly
 530 535 540
 5 Gln Thr Gly Pro Leu Pro Arg Ser Pro Leu Pro Gln Ser Pro Asn Pro
 545 550 555 560
 Gly Arg Arg His Gly Glu Glu Gly Gln Leu Gly Val Pro Thr Gly Glu
 565 570 575
 10 Leu Thr Ala Gly Ala Pro Glu Gly Pro Ala Leu Asp Thr Thr Gly Gln
 580 585 590
 Lys Ser Phe Leu Ser Ala Gly Tyr Leu Asn Glu Pro Phe Arg Ala Gln
 595 600 605
 15 Arg Ala Met Ser Val Val Ser Ile Met Thr Ser Val Ile Glu Glu Leu
 610 615 620
 Glu Glu Ser Lys Leu Lys Cys Pro Pro Cys Leu Ile Ser Phe Ala Gln
 625 630 635 640
 Lys Tyr Leu Ile Trp Glu Cys Cys Pro Lys Trp Arg Lys Phe Lys Met
 645 650 655
 25 Ala Leu Phe Glu Leu Val Thr Asp Pro Phe Ala Glu Leu Thr Ile Thr
 660 665 670
 Leu Cys Ile Val Val Asn Thr Val Phe Met Ala Met Glu His Tyr Pro
 675 680 685
 30 Met Thr Asp Ala Phe Asp Ala Met Leu Gln Ala Gly Asn Ile Val Phe
 690 695 700
 Thr Val Phe Phe Thr Met Glu Met Ala Phe Lys Ile Ile Ala Phe Asp
 705 710 715 720
 Pro Tyr Tyr Tyr Phe Gln Lys Lys Trp Asn Ile Phe Asp Cys Val Ile
 725 730 735
 40 Val Thr Val Ser Leu Leu Glu Leu Ser Ala Ser Lys Lys Gly Ser Leu
 740 745 750
 Ser Val Leu Arg Ser Leu Arg Leu Leu Arg Val Phe Lys Leu Ala Lys
 755 760 765
 45 Ser Trp Pro Thr Leu Asn Thr Leu Ile Lys Ile Ile Gly Asn Ser Val
 770 775 780
 Gly Ala Leu Gly Asn Leu Thr Phe Ile Leu Ala Ile Ile Val Phe Ile
 785 790 795 800
 Phe Ala Leu Val Gly Lys Gln Leu Leu Ser Glu Asp Tyr Gly Cys Arg
 805 810 815
 55 Lys Asp Gly Val Ser Val Trp Asn Gly Glu Lys Leu Arg Trp His Met
 820 825 830
 Cys Asp Phe Phe His Ser Phe Leu Val Val Phe Arg Ile Leu Cys Gly
 835 840 845
 60 Glu Trp Ile Glu Asn Met Trp Val Cys Met Glu Val Ser Gln Lys Ser
 850 855 860
 Ile Cys Leu Ile Leu Phe Leu Thr Val Met Val Leu Gly Asn Leu Val

-96-

	865		870		875		880
	Val Leu Asn Leu Phe	Ile Ala Leu Leu Leu Asn Ser Phe Ser Ala Asp					
		885		890			895
5	Asn Leu Thr Ala Pro Glu Asp Asp Gly Glu Val Asn Asn Leu Gln Leu						
		900		905			910
10	Ala Leu Ala Arg Ile Gln Val Leu Gly His Arg Ala Ser Arg Ala Ile						
		915		920			925
	Ala Ser Tyr Ile Ser Ser His Cys Arg Phe Arg Trp Pro Lys Val Glu						
		930		935			940
15	Thr Gln Leu Gly Met Lys Pro Pro Leu Thr Ser Ser Glu Ala Lys Asn						
			950		955		960
20	His Ile Ala Thr Asp Ala Val Ser Ala Ala Val Gly Asn Leu Thr Lys						
		965		970			975
	Pro Ala Leu Ser Ser Pro Lys Glu Asn His Gly Asp Phe Ile Thr Asp						
		980		985			990
25	Pro Asn Val Trp Val Ser Val Pro Ile Ala Glu Gly Glu Ser Asp Leu						
		995		1000			1005
	Asp Glu Leu Glu Glu Asp Met Glu Gln Ala Ser Gln Ser Ser Trp Gln						
		1010		1015			1020
30	Glu Glu Asp Pro Lys Gly Gln Gln Glu Gln Leu Pro Gln Val Gln Lys						
		1025		1030			1035
							1040
35	Cys Glu Asn His Gln Ala Ala Arg Ser Pro Ala Ser Met Met Ser Ser						
		1045		1050			1055
	Glu Asp Leu Ala Pro Tyr Leu Gly Glu Ser Trp Lys Arg Lys Asp Ser						
		1060		1065			1070
40	Pro Gln Val Pro Ala Glu Gly Val Asp Asp Thr Ser Ser Ser Glu Gly						
		1075		1080			1085
	Ser Thr Val Asp Cys Pro Asp Pro Glu Glu Ile Leu Arg Lys Ile Pro						
		1090		1095			1100
45	Glu Leu Ala Asp Asp Leu Asp Glu Pro Asp Asp Cys Phe Thr Glu Gly						
		1105		1110			1115
							1120
50	Cys Thr Arg Arg Cys Pro Cys Cys Asn Val Asn Thr Ser Lys Ser Pro						
		1125		1130			1135
	Trp Ala Thr Gly Trp Gln Val Arg Lys Thr Cys Tyr Arg Ile Val Glu						
		1140		1145			1150
55	His Ser Trp Phe Glu Ser Phe Ile Ile Phe Met Ile Leu Leu Ser Ser						
		1155		1160			1165
	Gly Ala Leu Ala Phe Glu Asp Asn Tyr Leu Glu Glu Lys Pro Arg Val						
		1170		1175			1180
60	Lys Ser Val Leu Glu Tyr Thr Asp Arg Val Phe Thr Phe Ile Phe Val						
		1185		1190			1195
							1200

	Phe	Glu	Met	Leu	Leu	Lys	Trp	Val	Ala	Tyr	Gly	Phe	Lys	Lys	Tyr	Phe	
																	1205 1210 1215
5	Thr	Asn	Ala	Trp	Cys	Trp	Leu	Asp	Phe	Leu	Ile	Val	Asn	Ile	Ser	Leu	
																	1220 1225 1230
	Thr	Ser	Leu	Ile	Ala	Lys	Ile	Leu	Glu	Tyr	Ser	Asp	Val	Ala	Ser	Ile	
																	1235 1240 1245
10	Lys	Ala	Leu	Arg	Thr	Leu	Arg	Ala	Leu	Arg	Pro	Leu	Arg	Ala	Leu	Ser	
																	1250 1255 1260
	Arg	Phe	Glu	Gly	Met	Arg	Val	Val	Val	Asp	Ala	Leu	Val	Gly	Ala	Ile	
																	1265 1270 1275 1280
15	Pro	Ser	Ile	Met	Asn	Val	Leu	Leu	Val	Cys	Leu	Ile	Phe	Trp	Leu	Ile	
																	1285 1290 1295
20	Phe	Ser	Ile	Met	Gly	Val	Asn	Leu	Phe	Ala	Gly	Lys	Phe	Ser	Lys	Cys	
																	1300 1305 1310
	Val	Asp	Thr	Arg	Asn	Asn	Pro	Phe	Ser	Asn	Val	Asn	Ser	Thr	Met	Val	
																	1315 1320 1325
25	Asn	Asn	Lys	Ser	Glu	Cys	His	Asn	Gln	Asn	Ser	Thr	Gly	His	Phe	Phe	
																	1330 1335 1340
	Trp	Val	Asn	Val	Lys	Val	Asn	Phe	Asp	Asn	Val	Ala	Met	Gly	Tyr	Leu	
																	1345 1350 1355 1360
30	Ala	Leu	Leu	Gln	Val	Ala	Thr	Phe	Lys	Gly	Trp	Met	Asp	Ile	Met	Tyr	
																	1365 1370 1375
	Ala	Ala	Val	Asp	Ser	Gly	Glu	Ile	Asn	Ser	Gln	Pro	Asn	Trp	Glu	Asn	
35																	1380 1385 1390
	Asn	Leu	Tyr	Met	Tyr	Leu	Tyr	Phe	Val	Val	Phe	Ile	Ile	Phe	Gly	Gly	
																	1395 1400 1405
40	Phe	Phe	Thr	Leu	Asn	Leu	Phe	Val	Gly	Val	Ile	Ile	Asp	Asn	Phe	Asn	
																	1410 1415 1420
	Gln	Gln	Lys	Lys	Lys	Leu	Gly	Gly	Gln	Asp	Ile	Phe	Met	Thr	Glu	Glu	
																	1425 1430 1435 1440
45	Gln	Lys	Lys	Tyr	Tyr	Asn	Ala	Met	Lys	Lys	Leu	Gly	Ser	Lys	Lys	Pro	
																	1445 1450 1455
	Gln	Lys	Pro	Ile	Pro	Arg	Pro	Leu	Asn	Lys	Tyr	Gln	Gly	Phe	Val	Phe	
50																	1460 1465 1470
	Asp	Ile	Val	Thr	Arg	Gln	Ala	Phe	Asp	Ile	Ile	Ile	Met	Val	Leu	Ile	
																	1475 1480 1485
55	Cys	Leu	Asn	Met	Ile	Thr	Met	Met	Val	Glu	Thr	Asp	Glu	Gln	Gly	Glu	
																	1490 1495 1500
	Glu	Lys	Thr	Lys	Val	Leu	Gly	Arg	Ile	Asn	Gln	Phe	Phe	Val			

-98-

Phe Thr Asn Gly Trp Asn Val Phe Asp Phe Ile Val Val Ile Leu Ser
 1540 1545 1550
 5 Ile Gly Ser Leu Leu Phe Ser Ala Ile Leu Lys Ser Leu Glu Asn Tyr
 1555 1560 1565
 Phe Ser Pro Thr Leu Phe Arg Val Ile Arg Leu Ala Arg Ile Gly Arg
 1570 1575 1580
 10 Ile Leu Arg Leu Ile Arg Ala Ala Lys Gly Ile Arg Thr Leu Leu Phe
 1585 1590 1595 1600
 Ala Leu Met Met Ser Leu Pro Ala Leu Phe Asn Ile Gly Leu Leu Leu
 1605 1610 1615
 15 Phe Leu Val Met Phe Ile Tyr Ser Ile Phe Gly Met Ala Ser Phe Ala
 1620 1625 1630
 Asn Val Val Asp Glu Ala Gly Ile Asp Asp Met Phe Asn Phe Lys Thr
 1635 1640 1645
 Phe Gly Asn Ser Met Leu Cys Leu Phe Gln Ile Thr Thr Ser Ala Gly
 1650 1655 1660
 25 Trp Asp Gly Leu Leu Ser Pro Ile Leu Asn Thr Gly Pro Pro Tyr Cys
 1665 1670 1675 1680
 Asp Pro Asn Leu Pro Asn Ser Asn Gly Ser Arg Gly Asn Cys Gly Ser
 1685 1690 1695
 30 Pro Ala Val Gly Ile Ile Phe Phe Thr Thr Tyr Ile Ile Ile Ser Phe
 1700 1705 1710
 Leu Ile Val Val Asn Met Tyr Ile Ala Val Ile Leu Glu Asn Phe Asn
 1715 1720 1725
 35 Val Ala Thr Glu Glu Ser Thr Glu Pro Leu Ser Glu Asp Asp Phe Asp
 1730 1735 1740
 40 Met Phe Tyr Glu Thr Trp Glu Lys Phe Asp Pro Glu Ala Thr Gln Phe
 1745 1750 1755 1760
 Ile Ala Phe Ser Ala Leu Ser Asp Phe Ala Asp Thr Leu Ser Gly Pro
 1765 1770 1775
 45 Leu Arg Ile Pro Lys Pro Asn Gln Asn Ile Leu Ile Gln Met Asp Leu
 1780 1785 1790
 Pro Leu Val Pro Gly Asp Lys Ile His Cys Leu Asp Ile Leu Phe Ala
 1795 1800 1805
 50 Phe Thr Lys Asn Val Leu Gly Glu Ser Gly Glu Leu Asp Ser Leu Lys
 1810 1815 1820
 55 Thr Asn Met Glu Glu Lys Phe Met Ala Thr Asn Leu Ser Lys Ala Ser
 1825 1830 1835 1840
 Tyr Glu Pro Ile Ala Thr Thr Leu Arg Trp Lys Gln Glu Asp Leu Ser
 1845 1850 1855
 60 Ala Thr Val Ile Gln Lys Ala Tyr Arg Ser Tyr Met Leu His Arg Ser
 1860 1865 1870

-99-

Leu Thr Leu Ser Asn Thr Leu His Val Pro Arg Ala Glu Glu Asp Gly
 1875 1880 1885
 5 Val Ser Leu Pro Gly Glu Gly Tyr Ser Thr Phe Met Ala Asn Ser Gly
 1890 1895 1900
 Leu Pro Asp Lys Ser Glu Thr Ala Ser Ala Thr Ser Phe Pro Pro Ser
 1905 1910 1915 1920
 10 Tyr Asp Ser Val Thr Arg Gly Leu Ser Asp Arg Ala Asn Ile Asn Pro
 1925 1930 1935
 Ser Ser Ser Met Gln Asn Glu Asp Glu Val Ala Ala Lys Glu Gly Asn
 1940 1945 1950
 15 Ser Pro Gly Pro Gln
 1955

20 (2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 25 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CAGCTTCGCT CAGAAGTATC T

21

35

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 22 base pairs
 (B) TYPE: nucleic acid
 40 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

50 TTCTCGCCGT TCCACACGGA GA

22

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 4 amino acids
 55 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

60

-100-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Phe Arg Leu Met
1

5

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: peptide

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Thr Gln Asp Phe Trp Glu Asn Leu Tyr
1 5

20

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

25

(ii) MOLECULE TYPE: peptide

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Thr Gln Asp Tyr Trp Glu Asn Leu Tyr
1 5

35

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

40

(ii) MOLECULE TYPE: peptide

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Thr Gln Asp Cys Trp Glu Arg Leu Tyr
1 5

50

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

55

(ii) MOLECULE TYPE: peptide

60

-101-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Thr Gln Asp Ser Trp Glu Arg Leu Tyr
1 5

5

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: peptide

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Thr Gln Asp Phe Trp Glu Arg Leu Tyr
1 5

20

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

25

(ii) MOLECULE TYPE: peptide

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Thr Gln Asp Ser Trp Glu Arg
1 5

35

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

40

(ii) MOLECULE TYPE: peptide

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Gly Ser Thr Asp Asp Asn Arg Ser Pro Gln Ser Asp Pro Tyr Asn
1 5 10 15

50

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

55

(ii) MOLECULE TYPE: peptide

60

RECEIVED
AUG 24 2000
TECH CENTER 1600/2900

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Ser Pro Lys Glu Asn His Gly Asp Phe Ile
1 5 10

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Pro Asn His Asn Gly Ser Arg Gly Asn
1 5

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Arg Leu Leu Arg Val Phe Lys Leu Ala Lys Ser Trp Pro Thr Leu
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GCTTGCTGCG GGTCTTCAAG C

21

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

-103-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Leu Arg Ala Leu Pro Leu Arg Ala Leu Ser Arg Phe Glu Gly
1 5 10

5

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

10

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

20 ATCGAGACAG AGCCCGCAGC G

21

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

25

- (A) LENGTH: 44 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

35

ACGGGTGCCG CAAGGACGGC GTCTCCGTGT GGAACGGCGA GAAG

44

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

40

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

50

GGCTATCCTT CCTCTTCCAG CTCTACCCA GGTATGGAGC CAGGT

45

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

55

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27: 21
TCCCGTACGC TGCAGCTCTT T
5
(2) INFORMATION FOR SEQ ID NO:28:
10 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
15 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28: 15
20 CCCGGGGAAG GCTAC
(2) INFORMATION FOR SEQ ID NO:29:
25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
30 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29: 15
35 GTCGACACCA GAAAT
(2) INFORMATION FOR SEQ ID NO:30:
40 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
45 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30: 30
50 GGATCCTCTA GAGTCGACCT GCAGAAGGAA

-105-

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 24 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

15 TGACGCAGGA CTCCTGGGAG CGCC

24

CLAIMS

1. A mammalian sensory neuron sodium channel protein, wherein the sodium channel is insensitive to tetrodotoxin.
2. The sodium channel protein of claim 1 wherein said protein is derived from dorsal root ganglia.
3. The sodium channel protein of claim 2 wherein the sodium channel protein is a rat protein.
4. The sodium channel protein of claim 2 wherein the sodium channel protein is a human protein.
5. The sodium channel protein of claim 3 wherein said protein comprises the amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 or SEQ ID NO: 8.
6. The sodium channel protein of claim 5 wherein said protein comprises the amino acid sequence of SEQ ID NO:2.
7. The sodium channel protein of claim 3 wherein said protein comprises the amino acid sequence encoded by the insert deposited in NCIMB deposit number 40744.
8. A nucleic acid sequence encoding the sodium channel protein of claims 1-7 or a complementary strand thereof.
9. The nucleic acid sequence of claim 8 wherein said nucleic acid sequence comprises the coding portion of the nucleic acid sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5 or SEQ ID NO: 7.
10. The nucleic acid sequence of claim 9 wherein said nucleic acid sequence comprises the coding portion of the nucleic acid sequence shown in SEQ ID NO:1.
11. The nucleic acid that hybridizes to strand of claim 8 or claim 10.
12. A nucleic acid sequence encoding rat dorsal root ganglia sodium channel protein which comprises the sequence of the coding portion of the insert deposited in NCIMB deposit number 40744 or a complementary strand thereof.
13. A vector comprising a nucleic acid sequence of claims 8-12.
14. A host cell transformed or transfected with a nucleic acid sequence of claims 8-12.

15. A method for identifying modulators of mammalian dorsal root ganglion sodium channel, which channel is insensitive to tetrodotoxin, comprising contacting a test compound with said channel and detecting the activity of said channel.
16. An antibody specific for the sodium channel protein of claim 1.
17. A nucleic acid sequence encoding the sodium channel protein of claims 1-7.
18. An expression vector comprising a nucleic acid sequence as defined in claim 12.
19. A host cell comprising an expression vector as defined in claim 18.
20. A method of making a sodium channel protein as defined in any one of claims 1 to 7 which comprises culture of a host cell as defined in claim 19 under conditions suitable for expression of the sodium channel protein and optionally purifying the expressed sodium channel protein.

Figure 1a

Nucleic acid and amino acid sequence of TTXi DRG sodium channel

```
tagcttgcttctgctaattgctacccaggccttttagacagagaacagatggcagatggag
1  -----+-----+-----+-----+-----+-----+
atcgaacgaagacgattacgatggggtccggaaatctgtctcttgtctaccgtctacctc

tttcttattgccatgcgcaaacgctgagcccacctcatgatcccggaaccccatgggtttc
61  -----+-----+-----+-----+-----+-----+
aaagaataacggtacgcggtttgcgactcgggtggagtactagggcctggggtaccaaaa

agtagacaacctgggctaagaagagatctccgaccttatagagcagcaaagagtgtaaat
121 -----+-----+-----+-----+-----+-----+
tcattctgttggaacccgattcttctctagaggctggaatatctcgtcgtttctcacattta

tcttccccaagaagaatgagaagATGGAGCTCCCCTTGCGTCCGTGGGAACCTACCAATT
181 -----+-----+-----+-----+-----+-----+
agaaggggttcttcttactcttcTACCTCGAGGGGAAACGCAGGCACCCTTGATGGTTAA

M E L P F A S V G T T N F

TCAGACGGTTCACCTCCAGAGTCACTGGCAGAGATCGAGAAGCAGATTGCTGCTCACC GGG
241 -----+-----+-----+-----+-----+-----+
AGTCTGCCAAGTGAGGTCTCAGTGACCGTCTCTAGCTCTTCGTCTAACGACGAGTG GCCC

R R F T P E S L A E I E K Q I A A H R A

CAGCCAAGAAGGCCAGAACCAAGCACAGAGGACAGGAGGACAAGGGCGAGAAGCCCAGGC
301 -----+-----+-----+-----+-----+-----+
GTCGGTCTTCCGGTCTTGGTTCGTGTCTCCTGTCTCCTGTTCCCGCTCTTCGGGTCCG

A K K A R T K H R G Q E D K G E K P R P

CTCAGCTGGACTTGAAAGACTGTAACCAGCTGCCCAAGTTCTATGGTGAGCTCCCAGCAG
361 -----+-----+-----+-----+-----+-----+
GAGTCGACCTGAACCTTTCTGACATTGGTCGACGGGTCAAGATACCACTCGAGGGTCGTC

Q L D L K D C N Q L P K F Y G E L P A E

AACTGGTCGGGGAGCCCCCTGGAGGACCTAGACCCTTTCTACAGCACACACCGGACATTCA
421 -----+-----+-----+-----+-----+-----+
TTGACCAGCCCCCTCGGGGACCTCCTGGATCTGGGAAAGATGTCGTGTGTGGCCTGTAAGT

L V G E P L E D L D P F Y S T H R T F M

TGGTGTGAATAAAAGCAGGACCATTTCCAGATTCAAGTCCACTTGGGCCCTGTGGCTCT
481 -----+-----+-----+-----+-----+-----+
ACCACAACCTATTTTTCGTCTGGTAAAGGTCTAAGTCACGGTGAACCCGGGACACCGAGA

V L N K S R T I S R F S A T W A L W L F
```

2/17

TCAGTCCCTTCAACCTGATCAGAAGAACAGCCATCAAAGTGTCTGTCCATTCTGGTTCT
541 -----+-----+-----+-----+-----+-----+-----+
AGTCAGGGAAGTTGGACTAGTCTTCTTGTCTGGTAGTTTCACAGACAGGTAAGGACCAAGA

S P F N L I R R T A I K V S V H S W F S

CCATATTCATCACCATCACTATTTTGGTCAACTGCGTGTGCATGACCCGAAC TGATCTTC
601 -----+-----+-----+-----+-----+-----+-----+
GGTATAAGTAGTGGTAGTGATAAAACCAGTTGACGCACACGTACTGGGCTTGACTAGAAG

I F I T I T I L V N C V C M T R T D L P

CAGAGAAAGTCGAGTACGTCTTCACTGTCAATTTACACCTTCGAGGCTCTGATTAAGATAC
661 -----+-----+-----+-----+-----+-----+-----+
GTCTCTTTCAGCTCATGCAGAAGTGACAGTAAATGTGGAAGCTCCGAGACTAATTCTATG

E K V E Y V F T V I Y T F E A L I K I L

TGGCAAGAGGGTTTTGTCTAAATGAGTTCACTTATCTTCGAGATCCGTGGAAC TGGCTGG
721 -----+-----+-----+-----+-----+-----+-----+
ACCGTTCTCCCAAACAGATTTACTCAAGTGAATAGAAGCTCTAGGCACCTTGACCGACC

A R G F C L N E F T Y L R D P W N W L D

ACTTCAGTGTCAATTACCTTGGCGTATGTGGGTGCAGCGATAGACCTCCGAGGAATCTCAG
781 -----+-----+-----+-----+-----+-----+-----+
TGAAGTCACAGTAATGGAACCGCATACACCCACGTCGCTATCTGGAGGCTCCTTAGAGTC

F S V I T L A Y V G A A I D L R G I S G -

GCCTGCGGACATTCCGAGTTCTCAGAGCCCTGAAAAC TTTCTGTGATCCCAGGACTGA
841 -----+-----+-----+-----+-----+-----+-----+
CGGACGCTGTAAAGGCTCAAGAGTCTCGGGACTTTTGACAAAGACACTAGGCTCCTGACT

L R T F R V L R A L K T V S V I P G L K -

AGGTCATCGTGGGAGCCCTGATCCACTCAGTGAGGAAGCTGGCCGACGTGACTATCCTCA
901 -----+-----+-----+-----+-----+-----+-----+
TCCAGTAGCACCCCTCGGGACTAGGTGAGTCACTCCTTCGACCGGCTGCACTGATAGGAGT

V I V G A L I H S V R K L A D V T I L T

CAGTCTTCTGCCTGAGCGTCTTCGCCTTGGTGGGCCTGCAGCTCTTTAAGGGGAACCTTA
961 -----+-----+-----+-----+-----+-----+-----+
GTCAGAAGACGGACTCGCAGAAGCGGAACCAACCCGACGTCGAGAAATTCCCTTGAAT

V F C L S V F A L V G L Q L F K G N L K

AGAACAAATGCATCAGGAACGGAACAGATCCCCACAAGGCTGACAACCTCTCATCTGAAA
1021 -----+-----+-----+-----+-----+-----+-----+
TCTTGTTTACGTAGTCCTTGCCTTGTCTAGGGGTGTTCCGACTGTTGGAGAGTAGACTTT

N K C I R N G T D P H K A D N L S S E M

TGGCAGAATACATCTTCATCAAGCCTGGTACTACGGATCCCTTACTGTGCGGCAATGGGT
1081 -----+-----+-----+-----+-----+-----+-----+
ACCGTCTTATGTAGAAGTAGTTCTGGACCATGATGCCTAGGGAATGACACGCCGTTACCCA

A E Y I F I K P G T T D P L L C G N G S

3/17

1141 CTGATGCTGGTCACTGCCCTGGAGGCTATGTCTGCCTGAAAACCTCTGACAACCCGGATT
-----+-----+-----+-----+-----+
GACTACGACCAGTGACGGGACCTCCGATACAGACGGACTTTTGAGGACTGTTGGGCCTAA
D A G H C P G G Y V C L K T P D N P D F

1201 TTAACCTACACCAGCTTTGATTCTTTGCGTGGGCATTCTCTCACTGTTCCGCCTCATGA
-----+-----+-----+-----+-----+
AATTGATGTGGTCGAAACTAAGGAAACGCACCCGTAAGGAGAGTGACAAGGCGGAGTACT
N Y T S F D S F A W A F L S L F R L M T

1261 CGCAGGACTCCTGGGAGCGCCTGTACCAGCAGACACTCCGGGCTTCTGGGAAAATGTACA
-----+-----+-----+-----+-----+
GCGTCCTGAGGACCCTCGCGGACATGGTCGTCTGTGAGGCCCGAAGACCCCTTTTACATGT
Q D S W E R L Y Q Q T L R A S G K M Y M

1321 TGGTCTTTTTTCGTGCTGGTTATTTTCCTTGGATCGTTCTACCTGGTCAATTTGATCTTGG
-----+-----+-----+-----+-----+
ACCAGAAAAAGCACGACCAATAAAAGGAACCTAGCAAGATGGACCAGTTAACTAGAACC
V F F V L V I F L G S F Y L V N L I L A

1381 CCGTGGTCACCATGGCGTATGAAGAGCAGAGCCAGGCAACAATTGCAGAAATCGAAGCCA
-----+-----+-----+-----+-----+
GGCACCAGTGGTACCGCATACTTCTCGTCTCGGTCCGTTGTTAACGTCTTTAGCTTCGGT
V V T M A Y E E Q S Q A T I A E I E A K

1441 AGGAAAAAAAGTTCCAGGAAGCCCTTGAGGTGCTGCAGAAGGAACAGGAGGTGCTGGCAG
-----+-----+-----+-----+-----+
TCCTTTTTTTCAAGGTCCTTCGGGAACCTCCACGACGTCTTCCTTGCTCTCCACGACCGTC
E K K F Q E A L E V L Q K E Q E V L A A

1501 CCCTGGGGATTGACACGACCTCGCTCCAGTCCCACAGTGGATCACCCCTAGCCTCCAAAA
-----+-----+-----+-----+-----+
GGGACCCCTAACTGTGCTGGAGCGAGGTACGGGTGTCACCTAGTGGGAATCGGAGGTTTT
L G I D T T S L Q S H S G S P L A S K N

1561 ACGCCAATGAGAGAAGACCCAGGGTGAAATCAAGGGTGTGAGAGGGCTCCACGGATGACA
-----+-----+-----+-----+-----+
TGCGGTTACTCTCTTCTGGGTCCCACTTTAGTTCCCACAGTCTCCCGAGGTGCCTACTGT
A N E R R P R V K S R V S E G S T D D N

1621 ACAGGTCACCCCAATCTGACCCTTACAACCAGCGCAGGATGTCTTTCTAGGCCTGTCTT
-----+-----+-----+-----+-----+
TGTCCAGTGGGGTTAGACTGGGAATGTTGGTCGCGTCCTACAGAAAGGATCCGGACAGAA
R S P Q S D P Y N Q R R M S F L G L S S

1681 CAGGAAGACGCAGGGCTAGCCACGGCAGTGTGTTCCACTTCCGAGCGCCCAGCCAAGACA
-----+-----+-----+-----+-----+
GTCCTTCTGCGTCCCGATCGGTGCCGTACACAAGGTGAAGGCTCGCGGGTCGGTTCTGT
G R R R A S H G S V F H F R A P S Q D I

4/17

1741 TCTCATTTTCCTGACGGGATCACCCCTGATGATGGGGTCTTTTCACGGAGACCAGGAAAGCC
-----+-----+-----+-----+-----+-----+
AGAGTAAAGGACTGCCCTAGTGGGGACTACTACCCAGAAAGTGCCTCTGGTCCTTTTCGG
S F P D G I T P D D G V F H G D Q E S R
GTCGAGGTTCCATATTGCTGGGCAGGGGTGCTGGGCAGACAGGTCCACTCCCCAGGAGCC
1801 -----+-----+-----+-----+-----+-----+
CAGCTCCAAGGTATAACGACCCGTCACGACCCGTCGTCCAGGTGAGGGGTCTTCGG
R G S I L L G R G A G Q T G P L P R S P
CACTGCCTCAGTCCCCCAACCCTGGCCGTAGACATGGAGAAGAGGGACAGCTCGGAGTGC
1861 -----+-----+-----+-----+-----+-----+
GTGACGGAGTCAGGGGGTTGGGACCGGCATCTGTACCTCTTCTCCCTGTCGAGCCTCACG
L P Q S P N P G R R H G E E G Q L G V P
CCACTGGTGAGCTTACCGCTGGAGCGCCTGAAGGCCCGGCACTCGACACTACAGGGCAGA
1921 -----+-----+-----+-----+-----+-----+
GGTGACCACTCGAATGGCGACCTCGCGGACTTCCGGGCCGTGAGCTGTGATGTCCCGTCT
T G E L T A G A P E G P A L D T T G Q K
AGAGCTTCCTGTCTGCGGGCTACTTGAACGAACCTTTCCGAGCACAGAGGGCCATGAGCG
1981 -----+-----+-----+-----+-----+-----+
TCTCGAAGGACAGACGCCCCGATGAACTTGCTTGGAAAGGCTCGTGTCTCCCGGTACTCGC
S F L S A G Y L N E P F R A Q R A M S V
TTGTCAAGTATCATGACTTCTGTCTGAGGAGCTTGAAGAGTCTAAGCTGAAGTGCCAC
2041 -----+-----+-----+-----+-----+-----+
AACAGTCATAGTACTGAAGACAGTAACTCCTCGAATCTCAGATTCTGACTTCACGGGTG
V S I M T S V I E E L E E S K L K C P P
CCTGCTTGATCAGCTTCGCTCAGAAGTATCTGATCTGGGAGTGCTGCCCCAAGTGGAGGA
2101 -----+-----+-----+-----+-----+-----+
GGACGAACTAGTCAAGCGAGTCTTCATAGACTAGACCCTCACGACGGGGTTCACCTCCT
C L I S F A Q K Y L I W E C C P K W R K
AGTTCAAGATGGCGCTGTTTCGAGCTGGTGACTGACCCCTTCGCAGAGCTTACCATCACCC
2161 -----+-----+-----+-----+-----+-----+
TCAAGTTCTACCGCGACAAGCTCGACCACTGACTGGGGAAGCGTCTCGAATGGTAGTGGG
F K M A L F E L V T D P F A E L T I T L
TCTGCATCGTGGTGAACACCGTCTTCATGGCCATGGAGCACTACCCCATGACCGATGCCT
2221 -----+-----+-----+-----+-----+-----+
AGACGTAGCACCCTTGTGGCAGAAGTACCGGTACCTCGTGATGGGGTACTGGCTACGGA
C I V V N T V F M A M E H Y P M T D A F
TCGATGCCATGCTTCAAGCCGGCAACATTGTCTTACCGTGTTTTTTCACAATGGAGATGG
2281 -----+-----+-----+-----+-----+-----+
AGCTACGGTACGAAGTTCCGGCCGTTGTAACAGAAGTGGCACAAAAGTGTACCTCTACC
D A M L Q A G N I V F T V F F T M E M A

5/17

2341 CCTTCAAGATCATTGCCCTTCGACCCCTACTATTACTTCCAGAAGAAGTGGAAATATCTTCG
-----+-----+-----+-----+-----+
GGAAGTTCTAGTAACGGAAGCTGGGGATGATAATGAAGGTCTTCTTCACCTTATAGAAGC

F K I I A F D P Y Y Y F Q K K W N I F D

2401 ACTGTGTCATCGTCACCGTGAGCCTTCTGGAGCTGAGTGCATCCAAGAAGGGCAGCCTGT
-----+-----+-----+-----+-----+
TGACACAGTAGCAGTGGCACTCGGAAGACCTCGACTCACGTAGGTTCTTCCCGTCGGACA

C V I V T V S L L E L S A S K K G S L S

2461 CTGTGCTCCGTTCTTACGCTTGCTGCGGGTCTTCAAGCTGGCCAAGTCTGGCCCACCC
-----+-----+-----+-----+-----+
GACACGAGGCAAGGAATGCGAACGACGCCCAGAAGTTCGACCGGTTACGACCGGGTGGG

V L R S L R L L R V F K L A K S W P T L

2521 TGAACACCCTCATCAAGATCATCGGGAAGTCACTGGGGGGCCCTGGGCAACCTGACCTTTA
-----+-----+-----+-----+-----+
ACTTGTGGGAGTAGTTCTAGTAGCCCTTGAGTCACCCCGGGACCCGTTGGACTGGAAT

N T L I K I I G N S V G A L G N L T F I

2581 TCCTGGCCATCATCGTCTTCATCTTCGCCCTGGTCGGAAGCAGCTTCTCTCAGAGGACT
-----+-----+-----+-----+-----+
AGGACCGGTAGTAGCAGAAGTAGAAGCGGGACCAGCCTTTCGTGAAGAGAGTCTCCTGA

L A I I V F I F A L V G K Q L L S E D Y

2641 ACGGGTGCCCAAGGACGGCGTCTCCGTGTGGAACGGCGAGAAGCTCCGCTGGCACATGT
-----+-----+-----+-----+-----+
TGCCACGGCGTTCCTGCCGAGAGGCACACCTTGCCGCTCTTCGAGGCGACCGTGTA

G C R K D G V S V W N G E K L R W H M C

2701 GTGACTTCTTCCATTCTTCTGCTGCTTCCGAATCCTCTGCGGGGAGTGGATCGAGA
-----+-----+-----+-----+-----+
CACTGAAGAAGGTAAGGAAGGACCAGCAGAAGGCTTAGGAGACGCCCTCACCTAGCTCT

D F F H S F L V V F R I L C G E W I E N

2761 ACATGTGGGTCTGCATGGAGGTCAGCCAGAAATCCATCTGCCTCATCTCTTCTTGA
-----+-----+-----+-----+-----+
TGTACACCCAGACGTACCTCCAGTCGGTCTTTAGGTAGACGGAGTAGGAGAAGAACTGAC

M W V C M E V S Q K S I C L I L F L T V

2821 TGATGGTGCTGGGCAACCTAGTGGTGCTCAACCTTTTCATCGCTTTACTGCTGAACCTCT
-----+-----+-----+-----+-----+
ACTACCACGACCCGTTGGATCACCACGAGTTGGAAGTAGCGAAATGACGACTTGAGGA

M V L G N L V V L N L F I A L L L N S F

2881 TCAGCGCGGACAACCTCACGGCTCCAGAGGATGACGGGGAGGTGAACAACCTTGCACTTAG
-----+-----+-----+-----+-----+
AGTCGCGCCTGTTGGAGTGCCGAGGTCTCCTACTGCCCCCTCACTTGTTGAACGTCAATC

S A D N L T A P E D D G E V N N L Q L A

6/17

2941 CACTGGCCAGGATCCAGGTACTTGGCCATCGGGCCAGCAGGGCCATCGCCAGTTACATCA
-----+-----+-----+-----+-----+-----+
GTGACCGGTCTAGGTCCATGAACCGGTAGCCCGGTCTCCCGGTAGCGGTCAATGTAGT

L A R I Q V L G H R A S R A I A S Y I S

3001 GCAGCCACTGCCGATTCCGCTGGCCCAAGGTGGAGACCCAGCTGGGCATGAAGCCCCAC
-----+-----+-----+-----+-----+-----+
CGTCGGTGACGGCTAAGGCGACCGGGTTCCACCTCTGGGTGACCCGTACTTCGGGGGTG

S H C R F R W P K V E T Q L G M K P P L

3061 TCACCAGCTCAGAGGCCAAGAACCACATTGCCACTGATGCTGTCAGTGCTGCAGTGGGGA
-----+-----+-----+-----+-----+-----+
AGTGGTCGAGTCTCCGGTTCTTGGTGTAACGGTGACTACGACAGTCACGACGTCACCCCT

T S S E A K N H I A T D A V S A A V G N

3121 ACCTGACAAAGCCAGCTCTCAGTAGCCCCAAGGAGAATCACGGGGACTTCATCACTGATC
-----+-----+-----+-----+-----+-----+
TGGACTGTTTCGGTTCGAGAGTCATCGGGGTTCTCTTAGTGCCCTGAAGTAGTGACTAG

L T K P A L S S P K E N H G D F I T D P

3181 CCAACGTGTGGGTCTCTGTGCCCATTGCTGAGGGGGAATCTGACCTCGACGAGCTCGAGG
-----+-----+-----+-----+-----+-----+
GGTTGCACACCCAGAGACACGGTAACGACTCCCCCTTAGACTGGAGCTGCTCGAGCTCC

N V W V S V P I A E G E S D L D E L E E

3241 AAGATATGGAGCAGGCTTCGAGAGCTCCTGGCAGGAAGAGGACCCCAAGGGACAGCAGG
-----+-----+-----+-----+-----+-----+
TTCTATACCTCGTCCGAAGCGTCTCGAGGACCGTCCTTCTCCTGGGGTTCCCTGTCTCGTCC

D M E Q A S Q S S W Q E E D P K G Q Q E

3301 AGCAGTTGCCACAAGTCCAAAAGTGTGAAAACCACCAGGCAGCCAGAAGCCCAGCCTCCA
-----+-----+-----+-----+-----+-----+
TCGTCAACGGTGTTCAGGTTTTTCACACTTTTGGTGGTCCGTCGGTCTTCGGGTTCGAGGT

Q L P Q V Q K C E N H Q A A R S P A S M

3361 TGATGTCTCTGAGGACCTGGCTCCATACCTGGGTGAGAGCTGGAAGAGGAAGGATAGCC
-----+-----+-----+-----+-----+-----+
ACTACAGGAGACTCCTGGACCGAGGTATGGACCCACTCTCGACCTTCTCCTTCCTATCGG

M S S E D L A P Y L G E S W K R K D S P

3421 CTCAGGTCCCTGCCGAGGGAGTGGATGACACGAGCTCCTCTGAGGGCAGCACGGTGGACT
-----+-----+-----+-----+-----+-----+
GAGTCCAGGGACGGCTCCCTCACCTACTGTGCTCGAGGAGACTCCCGTCGTGCCACCTGA

Q V P A E G V D D T S S S E G S T V D C

3481 GCGCGGACCCAGAGGAAATCCTGAGGAAGATCCCCGAGCTGGCAGATGACCTGGACGAGC
-----+-----+-----+-----+-----+-----+
CGGGCCTGGGTCTCCTTTAGGACTCCTTCTAGGGGCTCGACCGTCTACTGGACCTGCTCG

P D P E E I L R K I P E L A D D L D E P

7/17

3541 CCGATGACTGTTTCACAGAAGGCTGCACTCGCCGCTGTCCCTGCTGCAACGTGAATACTA
-----+-----+-----+-----+-----+-----+
GGCTACTGACAAAGTGCTTCCGACGTGAGCGGCGACAGGGACGACGTTGCACTTATGAT

D D C F T E G C T R R C P C C N V N T S

3601 GCAAGTCTCCTTGGGGCCACAGGCTGGCAGGTGCGCAAGACCTGCTACCGCATCGTGGAGC
-----+-----+-----+-----+-----+-----+
CGTTCAGAGGAACCCGGTGTCGACCGTCCACGCGTTCTGGACGATGGCGTAGCACCTCG

K S P W A T G W Q V R K T C Y R I V E H

3661 ACAGCTGGTTTGAGAGTTTCATCATCTTCATGATCCTGCTCAGCAGTGGAGCGCTGGCCT
-----+-----+-----+-----+-----+-----+
TGTCGACCAAACTCTCAAAGTAGTAGAAGTACTAGGACGAGTCGTCACCTCGCGACCGGA

S W F E S F I I F M I L L S S G A L A F

3721 TTGAGGATAACTACCTGGAAGAGAAACCCCGAGTGAAGTCCGTGCTGGAGTACACTGACC
-----+-----+-----+-----+-----+-----+
AACTCCTATTGATGGACCTTCTCTTTGGGGCTCACTTCAGGCACGACCTCATGTGACTGG

E D N Y L E E K P R V K S V L E Y T D R

3781 GAGTGTTACCTTCATCTTCGTCTTTGAGATGCTGCTCAAGTGGGTAGCCTATGGCTTCA
-----+-----+-----+-----+-----+-----+
CTCACAAGTGGAAAGTAGAAGCAGAACTCTACGACGAGTTCACCCATCGGATACCGAAGT

V F T F I F V F E M L L K W V A Y G F K

3841 AAAAGTATTTACCAATGCCTGGTGCTGGCTGGACTTCCTCATTTGTGAACATCTCCCTGA
-----+-----+-----+-----+-----+-----+
TTTTCATAAAGTGGTTACGGACCACGACCGACCTGAAGGAGTAACACTTGTAGAGGGACT

K Y F T N A W C W L D F L I V N I S L T

3901 CAAGCCTCATAGCGAAGATCCTTGAGTATTCGACGTGGCGTCCATCAAAGCCCTTCGGA
-----+-----+-----+-----+-----+-----+
GTTCCGAGTATCGCTTCTAGGAACTCATAAGGCTGCACCGCAGGTAGTTTCGGGAAGCCT

S L I A K I L E Y S D V A S I K A L R T

3961 CTCTCCGTGCCCTCCGACCGCTGCGGGCTCTGTCTCGATTGCAAGGCATGAGGGTAGTGG
-----+-----+-----+-----+-----+-----+
GAGAGGCACGGGAGGCTGGCGACGCCCCGAGACAGAGCTAAGCTTCCGTACTCCCATCACC

L R A L R P L R A L S R F E G M R V V V

4021 TGGATGCCCTCGTGGGCGCCATCCCCTCCATCATGAACGTCTCCTCGTCTGCCTCATCT
-----+-----+-----+-----+-----+-----+
ACCTACGGGAGCACCCGCGGTAGGGGAGGTAGTACTTGCAAGGAGGAGCAGACGGAGTAGA

D A L V G A I P S I M N V L L V C L I F

4081 TCTGGCTCATCTTCAGCATCATGGGCGTGAACCTCTTCGCCGGGAAATTTTGAAGTGCG
-----+-----+-----+-----+-----+-----+
AGACCGAGTAGAAGTCGTAGTACCCGCACTGGAGAAGCGGCCCTTTAAAAGCTTCACGC

W L I F S I M G V N L F A G K F S K C V

8/17

4141 TCGACAGCAGAAATAACCCATTTTCCAACGTGAATTTCGACGATGGTGAATAACAAGTCCG
-----+-----+-----+-----+-----+-----+
AGCTGTGGTCTTTATTGGGTAAAAGGTTGCACTTAAGCTGCTACCACTTATTGTTTCAGGC

D T R N N P F S N V N S T M V N N K S E

4201 AGTGTCACAATCAAAACAGCACCGGCCACTTCTTCTGGGTCAACGTCAAAGTCAACTTCG
-----+-----+-----+-----+-----+-----+
TCACAGTGTTAGTTTTGTCTGTGGCCGGTGAAGAAGACCCAGTTGCAGTTTCAGTTGAAGC

C H N Q N S T G H F F W V N V K V N F D

4261 ACAACGTCGCTATGGGCTACCTCGCACTTCTTCAGGTGGCAACCTTCAAAGGCTGGATGG
-----+-----+-----+-----+-----+-----+
TGTTGCAGCGATACCCGATGGAGCGTGAAGAAGTCCACCGTTGGAAGTTTCCGACCTACC

N V A M G Y L A L L Q V A T F K G W M D

4321 ACATAATGTATGCAGCTGTTGATTCCGGAGAGATCAACAGTCAGCCTAACTGGGAGAACA
-----+-----+-----+-----+-----+-----+
TGTATTACATACGTCGACAACTAAGGCCTCTCTAGTTGTCTAGTCGGATTGACCCTCTTGT

I M Y A A V D S G E I N S Q P N W E N N

4381 ACTTGTACATGTACCTGTACTTCGTCGTTTTCATCATTTTCGGTGGCTTCTTCACGCTGA
-----+-----+-----+-----+-----+-----+
TGAACATGTACATGGACATGAAGCAGCAAAAGTAGTAAAAGCCACCGAAGAAGTGGCGACT

L Y M Y L Y F V V F I I F G G F F T L N

4441 ATCTCTTTGTTGGGGTCATAATCGACAACCTTCAACCAACAGAAAAAAGCTAGGAGGCC
-----+-----+-----+-----+-----+-----+
TAGAGAAACAACCCAGTATTAGCTGTTGAAGTTGGTTGTCTTTTTTTTCGATCCTCCGG

L F V G V I I D N F N Q Q K K K L G G Q

4501 AGGACATCTTCATGACAGAAGAGCAGAAGAAGTACTACAATGCCATGAAGAAGCTGGGCT
-----+-----+-----+-----+-----+-----+
TCCTGTAGAAGTACTGTCTTCTCGTCTTCTTCATGATGTTACGGTACTTCTTCGACCCGA

D I F M T E E Q K K Y Y N A M K K L G S

4561 CCAAGAAACCCCAAGCCCATCCACGGCCCTGAATAAGTACCAAGGCTTCGTGTTTG
-----+-----+-----+-----+-----+-----+
GGTCTTTGGGGTCTTCGGGTAGGGTGCCGGGGACTTATTCATGGTTCCGAAGCACAAAC

K K P Q K P I P R P L N K Y Q G F V F D

4621 ACATCGTGACCAGGCAAGCCTTTGACATCATCATGTTCTCATCTGCCTCAACATGA
-----+-----+-----+-----+-----+-----+
TGTAGCACTGGTCCGTTTCGGAACTGTAGTAGTAGTACCAAGAGTAGACGGAGTTGTACT

I V T R Q A F D I I I M V L I C L N M I

4681 TCACCATGATGGTGGAGACCGACGAGCAGGGCGAGGAGAAGACGAAGGTTCTGGGCAGAA
-----+-----+-----+-----+-----+-----+
AGTGGTACTACCACCTCTGGCTGCTCGTCCCGCTCCTCTTCTGCTTCCAAGACCCGTCTT

T M M V E T D E Q G E E K T K V L G R I